

**CELL LINES AND HOST NUCLEIC ACID SEQUENCES RELATED TO
INFECTIOUS DISEASE**

CROSS REFERENCE TO RELATED APPLICATIONS

5 This application claims the benefit of U.S. Provisional Application Nos. 60/427,464 filed November 18, 2002 and 60/482,604 filed June 25, 2003, both herein incorporated by reference.

FIELD

10 The present disclosure relates to host nucleic acid sequences, and proteins encoded by these sequences, that are involved in viral infection or are otherwise associated with the life cycle of a virus. Decreasing or inhibiting the interaction of these host sequences with a viral sequence can be used to decrease or inhibit infection by the virus.

BACKGROUND

15 Infectious diseases affect the health of people and animals around the world, causing serious illness and death. Public health efforts have focused on behavioral modification and other public health efforts to reduce the incidences of infection, while treatment regimens for these diseases have focused on pharmaceuticals, such as antibiotics and anti-viral medications. However, educating people about modifying behavior can be difficult, and that approach alone rarely can significantly diminish the incidence of infection. Furthermore, modifying the behavior of domestic or wild animals would not result in diminished infections. Stopping the spread of infections in an animal population typically involves wholesale slaughter. Few vaccines are available or wholly effective, and they tend to be specific for particular conditions.

20 The rate of HIV (human immunodeficiency virus) infection is increasing. HIV and its associated acquired immune deficiency syndrome (AIDS) accounted for approximately 5% of all deaths in the United States in the year 2000, while over 313,000 persons were reported to be living with AIDS in that same year. Centers for Disease Control and Prevention, *HIV/AIDS Surveillance Supplemental Report*, 8(1):1-22 (2002). These increasing infection rates have occurred, even though the mode of HIV infection has been known for almost 20 years, and educational programs around the world have promoted behavioral modifications meant to reduce HIV infection. Incidence and death rates due to HIV disease have been decreasing since the mid-90's, in part due to aggressive antiviral therapies, which frequently have toxic side effects and strict dosage schedules. However, even with treatment, the patient is not cured of the disease, and to date, no effective vaccine therapy has been found.

25 In other diseases, such as infection by the Ebola virus, not only are treatments limited, but containment or prevention of infections is difficult because the life cycle of the virus is not well known. The natural reservoir for the Ebola virus, that is the place or population in nature where the virus resides between human outbreaks, has not yet been identified.

Additionally, different viral strains can rapidly evolve in response to drug usage, producing drug-resistant strains. For example, strains of the influenza virus resistant to amantadine and rimantadine have recently arisen. A recent study of 80 newly-infected people conducted by the AIDS Research Center at Rockefeller University in New York, found that as many as 16.3% of these 5 individuals had strains of HIV associated with resistance to some treatments, and 3.8% appeared to be resistant to several currently available anti-HIV drugs. Thus, a need exists for alternative treatments for infectious disease and methods of designing new drugs to combat infectious disease.

SUMMARY

10 Several host nucleic acid sequences involved in viral infection have been identified using gene trap methods. The identification of these host sequences and their encoded products permits the identification of sequences that can be targeted for therapeutic intervention.

The disclosed host sequences (including the target sequences associated with SEQ ID NOS: 1-227, 229, and 231, and the proteins encoded thereby (such as SEQ ID NOS: 228, 230, and 232), as 15 well as variants, fusions, and fragments thereof that retain the appropriate biological activity) can mediate infection, and in some examples these host nucleic acids are required for infection. For example, the host nucleic acid can encode a cellular receptor or ligand or a fragment thereof that is recognized by a virus, such as the T-cell V-D-J beta 2.1 chain. In another example, the host nucleic acid encodes an enzyme that mediates viral infection, such as the β -chimerin rho-GTPase (referred to 20 herein as β -chimerin). In another example, the host nucleic acid encodes a Ras oncogene family member such as Rab9. It is demonstrated herein that Rab9 is a host protein involved in infection by pathogens (such as viruses and bacteria) that use similar pathways for morphogenesis of infectious particles. In particular examples, Rab9 is involved in infection by pathogens (such as viruses and bacteria) that utilize lipid rafts. Thus, for example, interfering with the interaction between the 25 disclosed host proteins and a viral or pathogen protein, for example by disrupting the expression of the host nucleic acid within a host cell, or by administering an agent that decreases binding between a host protein and a viral protein, can inhibit, or even prevent, infection of that host cell by the associated virus. Moreover, the identification of particular host enzymes or other host proteins involved in infection provides a method for developing new therapies targeted at inhibiting infection, 30 at the protein or nucleic acid level.

In some examples, the nucleic acid itself mediates viral infection. For example, the nucleotide sequence of a host nucleic acid in the host genome can be recognized by the virus during integration of the viral genome into the host genome. The identification of nucleic acid sequences that are involved in the pathogenesis of infection therefore provides an important tool for interfering 35 with infection.

This genomics-based discovery of nucleic acids and proteins involved in, or even required for, infection provides a new paradigm for identifying and validating various aspects of infectious disease, including assessing individual or population resistance to infection and finding novel

diagnostic and drug targets for infectious disease and altering the nucleotide sequence of the host nucleic acid.

Based on the identification of several host nucleic acid and protein sequences involved in viral infection, provided herein are methods for decreasing infection of a host cell by a virus, such as HIV, Ebola, or influenza A, or treating such a viral infection, by interfering with the activity or expression of one or more host proteins shown in Table 1 (including the target sequences associated with any of SEQ ID NOS: SEQ ID NOS: 1-232, as well as variants, fragments, and fusions thereof), such as at least two host proteins, or at least three host proteins. Also provided are methods for identifying agents that can decrease viral infection of a host cell, such as infection by HIV, Ebola, or influenza A. In addition, cells and non-human mammals are provided that have decreased susceptibility to viral infection, such as HIV, Ebola, or influenza A infection.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 is a schematic illustration of the U3neoSV1 retroviral vector, which is capable of isolating the nucleic acids described herein using the gene-trap method.

FIG. 2 is a schematic illustration of the gene-trap method.

FIG. 3 is a schematic illustration of one method of identifying host genes described herein.

FIG. 4 is a flow chart illustrating a method for isolating cells resistant to HIV infection, including HIV-1 and HIV-2 infection.

FIG. 5 is a bar graph showing the relative amount of p24 in HIV-infected cells in the presence of various siRNAs. CHN (β -chimerin); KOX (similar to KOX4 (LOC131880) and LOC166140); RBB (retinoblastoma binding protein 1); RAB (Rab9); KIAA1259; F3 (tissue factor 3; thromboplastin); AXL (AXL receptor tyrosine kinase); Msleb (mammalian selenium binding protein).

FIG. 6 is a schematic drawing showing a model of Rab9 involvement in lipid raft formation.

SEQUENCE LISTING

The nucleotide sequences of the nucleic acids described herein are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. Additionally, the nucleic acid sequences shown in SEQ ID NOS: 1-226 inherently disclose the corresponding polypeptide sequences of coding sequences (resulting translations of the nucleotide sequences), even when those polypeptide sequences are not explicitly provided herein.

SEQ ID NO: 1 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18E8, entire insert. The human homolog is the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor V beta chain (T-cell receptor beta). Further information on the T-cell receptor V beta chain can be found in WO 01/23409, WO 01/55302, WO 01/57182, and WO 01/94629.

SEQ ID NO: 2 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18BA, distal end. The human homolog is the (-) strand of GenBank Accession No. AC104597.3, T-cell receptor V beta chain.

5 SEQ ID NO: 3 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18BA, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 4 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18BE, distal end. The human homolog is the (+) strand of GenBank Accession No. AC00616.7, T-cell receptor beta.

10 SEQ ID NO: 5 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18BE, middle of insert. The human homolog is the (-) strand of GenBank Accession No. AC104597.3, T-cell receptor beta.

15 SEQ ID NO: 6 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18BE, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 7 is a nucleic acid sequence associated with viral, such as HIV, infection which corresponds to the sequence identified as Nucleotide Sequence 18E6, proximal end. The human homolog is the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

20 SEQ ID NO: 8 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E21, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 9 is a nucleic acid sequence associated with viral, such as HIV, infection which corresponds to the sequence identified as Nucleotide Sequence 2E22, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC099395.2, T-cell receptor beta.

25 SEQ ID NO: 10 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B13, proximal end. The human homolog is the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 11 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B14, proximal end. The human homolog is the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

30 SEQ ID NO: 12 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B15, distal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 13 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B15, proximal end. The human homolog is the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 14 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B16, proximal end. The human homolog is the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 15 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E23, distal end. The human homolog is the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

5 SEQ ID NO: 16 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E23, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 17 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E24, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

10 SEQ ID NO: 18 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E25, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

15 SEQ ID NO: 19 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E26, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 20 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18BD, proximal end. The human homolog is the (+) strand of GenBank Accession No. M16834.1, T-cell receptor V-D-J-beta 2.1 chain (described in WO 02/057414 and Reynolds *et al.*, *Cell* 50(1):107-17, 1987).

20 SEQ ID NO: 21 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18E7, distal end. The human homolog is the (-) strand of GenBank Accession No. AC004593.1 including beta-chimaerin rho GTPase (CHN2) (for example see WO 01/12659).

25 SEQ ID NO: 22 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 18E7, proximal end. The human homologs are the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta; and the (+) strand of GenBank Accession No. AC004593.1 including beta-chimaerin (CHN2).

30 SEQ ID NO: 23 is a nucleic acid sequence associated with viral, such as HIV and influenza A, infection, and is the clone identified as Nucleotide Sequence 18E6, distal end. The human homolog is the (+) strand of GenBank Accession No. AL049699.8, including malic enzyme 1 (ME1) NADP(+)-dependent cytosolic. Further information on this gene can be found in WO 01/55301 and WO 01/53312.

35 SEQ ID NO: 24 is a nucleic acid sequence associated with viral, such as HIV and influenza A, infection, and is the clone identified as Nucleotide Sequence 18BD, distal end. The human homolog is the (+) strand of GenBank Accession No. AC123903.1, including hypothetical protein XP_174419.

SEQ ID NO: 25 is a nucleic acid sequence associated with viral, such as HIV and influenza A, infection, and is the clone identified as Nucleotide Sequence 18E9, distal end. The human

homolog is the (+) strand of GenBank Accession No. AC096736.3, a region of chromosome 4q31.3-32.

SEQ ID NO: 26 is a nucleic acid sequence associated with viral, such as HIV and influenza A, infection, and is the clone identified as Nucleotide Sequence 18E9, middle of insert. The human homolog is the (+) strand of GenBank Accession No. AC096736.3, a region of chromosome 4q31.3-32.

SEQ ID NO: 27 is a nucleic acid sequence associated with viral, such as HIV and influenza A, infection, and is the clone identified as Nucleotide Sequence 18E9, proximal end. The human homologs are the (-) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta; and (-) strand of GenBank Accession No. AC096736.3, a region of chromosome 4q31.3-32.

SEQ ID NO: 28 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E21, distal end. The human homolog is the (-) strand of GenBank Accession No. M26920.1, alpha satellite DNA.

SEQ ID NO: 29 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E22, distal end. The human homologs are the (+) strand of GenBank Accession No. AP004369.3, including LOC253788 (and neighboring similar to RIKEN cDNA 1700001L23 (LOC219938)); and the (+) strand of GenBank Accession No. AC093117.2, between coagulation factor III, thromboplastin, tissue factor (F3) and LOC91759.

SEQ ID NO: 30 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B13, distal end. The human homolog is the (-) strand of GenBank Accession No. AC092043.2, between similar to zinc finger protein 7 KOX4 (LOC131880) and LOC166140.

SEQ ID NO: 31 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B14, distal end. The human homologs are the (-) strand of GenBank Accession No. AL136963.17, between LOC222474 and similar to Rho guanine nucleotide exchange factor 4, isoform a, APC-stimulated guanine nucleotide exchange factor (LOC221178); and the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

SEQ ID NO: 32 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2B16, distal end. The human homolog is the (-) strand of GenBank Accession No. AL133293.28, between ribosomal protein L7A-like 4 (RPL7AL4) and v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian) (SRC).

SEQ ID NO: 33 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E24, distal end. The human homolog is the (-) strand of GenBank Accession No. AL161417.17, KIAA0564.

SEQ ID NO: 34 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E25, distal end. The human homologs are the (-) strand of GenBank Accession No. Z12006.1, alpha satellite DNA; and the (+) and (-) strands of GenBank Accession No. AC093577.2, M96 protein.

SEQ ID NO: 35 is a nucleic acid sequence associated with viral, such as HIV, infection, and is the clone identified as Nucleotide Sequence 2E26, distal end. The human homologs are the (-) strand of GenBank Accession No. Z78022.1, hypothetical protein similar to G proteins, especially RAP-2A (LOC57826); and the (+) strand of GenBank Accession No. AL136220.14, between LOC161005 and osteoblast specific factor 2 (fasciclin I-like; OSF-2).

SEQ ID NO: 36 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B3B1, distal end. The canine homolog is the (+) and (-) strand portions of GenBank Accession No. AJ012166.1, *Canis familiaris* TCTA gene, AMT gene, DAG1 gene, and BSN gene.

SEQ ID NO: 37 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B5B5, distal end. The canine homolog is the (+) and (-) strand portions of GenBank Accession No. AJ012166.1, *Canis familiaris* TCTA gene, AMT gene, DAG1 gene, and BSN gene.

SEQ ID NO: 38 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1B1, distal end. The human homolog is the (+) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 39 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1B2, distal end. The human homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 40 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1B4, distal end. The human homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 41 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1B5, distal end. The human homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 42 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1B6, distal end. The human homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 43 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1E3, entire insert. The human homolog is the (+) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 44 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1E5, proximal end. The human

homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 45 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6B1, entire insert. The human
5 homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 46 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E1, distal end. The human homolog
10 is the (+) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

SEQ ID NO: 47 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

15 SEQ ID NO: 48 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E3, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC117507.5, including LIM domain containing preferred translocation partner in lipoma (LPP).

20 SEQ ID NO: 49 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AC104036.8, between LOC253121 and hyaluronan synthase 2 (HAS2).

SEQ ID NO: 50 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1E1, proximal end. The human
25 homolog is the (-) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3 (see WO 01/57270, WO 01/57271, WO 01/57273, WO 01/57274, WO 01/57275, WO 01/57276, WO 01/57277, WO 01/57278, or Tatarelli *et al.*, *Genomics* 68(1):1-12, 2000).

30 SEQ ID NO: 51 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1E4, distal end. The human homolog is the (+) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3.

SEQ ID NO: 52 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B1E4, proximal end. The human homolog is the (-) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3.

35 SEQ ID NO: 53 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E2, distal end. The human homolog is the (+) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3.

SEQ ID NO: 54 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E2, proximal end. The human homolog is the (-) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3.

SEQ ID NO: 55 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E4, proximal end. The human homolog is the (-) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3.

5 SEQ ID NO: 56 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B7E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3.

SEQ ID NO: 57 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B7E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. AF260225.1, Testin 2 and Testin 3.

10 SEQ ID NO: 58 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B5E2, distal end. The human homolog is the (-) strand of GenBank Accession No. AL133230.25, PTPN1 gene for protein tyrosine phosphatase, non-receptor type 1 (see Watanabe *et al.*, *Jpn. J. Cancer Res.* 93:1114-22, 2002).

15 SEQ ID NO: 59 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B5E2, middle of insert. The human homolog is the (-) strand of GenBank Accession No. AL133230.25, PTPN1 gene for protein tyrosine phosphatase, non-receptor type 1.

20 SEQ ID NO: 60 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B3E11, distal end. The human homolog is the (+) strand of GenBank Accession No. AL445675.9, between LOC149360 and LOC253961.

SEQ ID NO: 61 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B3E11, proximal end. The human homolog is the (-) strand of GenBank Accession No. AL391986.12, between KIAA1560 and Tectorin beta (TECTB).

25 SEQ ID NO: 62 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E3, distal end. The human homolog is the (-) strand of GenBank Accession No. AC016826.9, including Cadherin related 23 (CDH23).

30 SEQ ID NO: 63 is a nucleic acid sequence associated with viral, such as influenza A, infection, and is the clone identified as Nucleotide Sequence B6E4, distal end. The human homolog is the (+) strand of GenBank Accession No. AL357372.12, Myeloid/lymphoma or mixed lineage leukemia, translocated to 10 (MMLT10).

35 SEQ ID NO: 64 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence ZV1-1B5, distal end. The human homolog is the (-) strand of GenBank Accession No. AL355802.13, between exportin 5 (XPO5) and DNA polymerase eta (POLH).

SEQ ID NO: 65 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence ZV1-1B5, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL355802.13, between XPO5 and POLH.

SEQ ID NO: 66 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence ZV1-1E, proximal end. The human homolog is the (-) strand of GenBank Accession No. AL355802.13, between XPO5 and POLH.

SEQ ID NO: 67 is a nucleic acid sequence associated with viral, such as Ebola, infection,
5 and is the clone identified as Nucleotide Sequence MV1-2E1, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL135744.4, including heterogenous nuclear riboprotein C (C1/C2) (HNRPC).

SEQ ID NO: 68 is a nucleic acid sequence associated with viral, such as Ebola, infection,
and is the clone identified as Nucleotide Sequence MV1-2E5, proximal end. The human homolog is
10 the (+) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 69 is a nucleic acid sequence associated with viral, such as Ebola, infection,
and is the clone identified as Nucleotide Sequence MV1-2E6, proximal end. The human homolog is
the (+) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 70 is a nucleic acid sequence associated with viral, such as Ebola, infection,
15 and is the clone identified as Nucleotide Sequence MV1-2B2, proximal end. The human homolog is
the (-) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 71 is a nucleic acid sequence associated with viral, such as Ebola, infection,
and is the clone identified as Nucleotide Sequence MV1-2B13, proximal end. The human homolog is
the (-) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 72 is a nucleic acid sequence associated with viral, such as Ebola, infection,
20 and is the clone identified as Nucleotide Sequence MV1-2B14, proximal end. The human homolog is
the (-) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 73 is a nucleic acid sequence associated with viral, such as Ebola, infection,
25 and is the clone identified as Nucleotide Sequence MV1-2B21, proximal end. The human homolog is
the (-) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 74 is a nucleic acid sequence associated with viral, such as Ebola, infection,
and is the clone identified as Nucleotide Sequence MV1-2B25, proximal end. The human homolog is
the (-) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 75 is a nucleic acid sequence associated with viral, such as Ebola, infection,
30 and is the clone identified as Nucleotide Sequence MV1-2B35, proximal end. The human homolog is
the (-) strand of GenBank Accession No. AL135744.4, including HNRPC.

SEQ ID NO: 76 is a nucleic acid sequence associated with viral, such as Ebola, infection,
and is the clone identified as Nucleotide Sequence MV1-2E5, distal end. The human homolog is the
35 (+) and (-) strands of GenBank Accession No. AL050324.5, including alpha-endosulfine pseudogene
(ENSAP) and LOC128741.

SEQ ID NO: 77 is a nucleic acid sequence associated with viral, such as Ebola, infection,
and is the clone identified as Nucleotide Sequence MV1-2E6, distal end. The human homolog is the
(+)
(+) strand of GenBank Accession No. AC017060.7, including LOC222888.

SEQ ID NO: 78 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV1-2B13, distal end. The human homolog is the (+) strand of GenBank Accession No. AL161731.20, between LOC138421 and zinc finger protein 297B (ZNF297B).

5 SEQ ID NO: 79 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV1-2B14, distal end. The human homolog is the (-) strand of GenBank Accession No. AC012366.10, including sideroflexin 5 (SFXN5).

10 SEQ ID NO: 80 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV1-2B35, distal end. The human homolog is the (+) strand of GenBank Accession No. AL645504.10, including importin 9 (FLJ10402).

15 SEQ ID NO: 81 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence GV1-1B1, distal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

20 SEQ ID NO: 82 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence GV1-1B1, proximal end. The human homolog is the (+) strand of GenBank Accession No. NG_001333.1, T-cell receptor beta.

25 SEQ ID NO: 83 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-B1, distal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

30 SEQ ID NO: 84 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E2, distal end. The human homolog is the (+) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

35 SEQ ID NO: 85 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E2, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

SEQ ID NO: 86 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E3, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

40 SEQ ID NO: 87 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E4, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

45 SEQ ID NO: 88 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E5, distal end. The human homolog is the (+) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

50 SEQ ID NO: 89 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E5, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

SEQ ID NO: 90 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-B1, distal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

5 SEQ ID NO: 91 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E2, distal end. The human homolog is the (+) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

SEQ ID NO: 92 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E2, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

10 SEQ ID NO: 93 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E4, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

15 SEQ ID NO: 94 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-B1, distal end. The human homolog is the (+) and (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

SEQ ID NO: 95 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

20 SEQ ID NO: 96 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

SEQ ID NO: 97 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E4, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

25 SEQ ID NO: 98 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E5, distal end. The human homolog is the (+) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

SEQ ID NO: 99 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E5, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, adjacent to LOC135952.

30 SEQ ID NO: 100 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AC021753.7, hypothetical protein KIAA1259.

35 SEQ ID NO: 101 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC021753.7, hypothetical protein KIAA1259.

SEQ ID NO: 102 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E3, distal end. The human homolog is the

(+) and (-) strands of GenBank Accession No. AC107081.5, copper metabolism gene (MURR1) and chaperonin containing TCP1, subunit 4 (CCT4).

SEQ ID NO: 103 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E4, distal end. The human homolog is the (-) strand of GenBank Accession No. AC099785.2, hypothetical protein FLJ40773 and similar to ribosomal protein L24-like (LOC149360).

SEQ ID NO: 104 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E4, distal end. The human homolog is the (+) strand of GenBank Accession No. AF260225.1, Testin 2 and 3 (TES).

SEQ ID NO: 105 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV2-E4, proximal end. The human homolog is the (-) strand of GenBank Accession No. AF260225.1, Testin 2 and 3 (TES).

SEQ ID NO: 106 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E7, distal end. The human homolog is the (+) strand of GenBank Accession No. AF260225.1, Testin 2 and 3 (TES).

SEQ ID NO: 107 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E7, proximal end. The human homolog is the (-) strand of GenBank Accession No. AF260225.1, Testin 2 and 3 (TES).

SEQ ID NO: 108 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-B2, distal end. The human homolog is the (+) and (-) strands of GenBank Accession No. AC105934.2, polybromo 1 (PB1).

SEQ ID NO: 109 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-B4, distal end. The human homolog is the (+) strand of GenBank Accession No. AC022506.38, between DNA damage inducible transcript 3 (DDIT3) and KIAA1887.

SEQ ID NO: 110 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-B5, distal end. The human homolog is the (-) strand of GenBank Accession No. AL157834.12, PDZ and LIM domain 1 (elfin) (PDLIM1).

SEQ ID NO: 111 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AL110115.38, LOC284803.

SEQ ID NO: 112 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. AL110115.38, signal peptide peptidase (HM13) and LOC284803.

SEQ ID NO: 113 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E3, distal end. The human homolog is the (-) strand of GenBank Accession No. AL117341.26, containing PRO0097 and adjacent to FLJ31958.

SEQ ID NO: 114 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E3, proximal end. The human homolog is the (-) strand of GenBank Accession No. AP002076.3, small inducible cytokine E, member 1 (endothelial monocyte-activating) (SCYE1).

5 SEQ ID NO: 115 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E6, distal end. The human homolog is the (+) strand of GenBank Accession No. AP002076.3, containing SCYE1.

10 SEQ ID NO: 116 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E6, proximal end. The human homolog is the (-) strand of GenBank Accession No. AP002076.3, containing SCYE1.

SEQ ID NO: 117 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E4, distal end. The human homolog is the (+) and (-) strands of GenBank Accession No. AC132812.9, between E3 ubiquitin ligase (SMURF2) and hypothetical protein MGC40489.

15 SEQ ID NO: 118 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E5, distal end. The human homolog is the (+) strand of GenBank Accession No. AC079383.17, Ras oncogene family member Rab9.

20 SEQ ID NO: 119 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV3-E5, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC079383.17, Ras oncogene family member Rab9.

SEQ ID NO: 120 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E2, distal end. The human homolog is the (-) strand of GenBank Accession No. AL132989.5, between PRO1617 and retinoblastoma binding protein 1 (RBBP1).

25 SEQ ID NO: 121 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E2, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL132989.5, RBBP1.

30 SEQ ID NO: 122 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E3, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL132989.5, retinoblastoma binding protein 1 (RBBP1).

SEQ ID NO: 123 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E3, distal end. The human homolog is the (+) and (-) strands of GenBank Accession No. AC096669.1, a region of chromosome 2q12.

35 SEQ ID NO: 124 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV6-E4, distal end. The human homolog is the (-) strands of GenBank Accession No. AF196968.4, elongation factor for selenoprotein translation (SELB).

SEQ ID NO: 125 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-B1, distal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

SEQ ID NO: 126 is a nucleic acid sequence associated with viral, such as Ebola, infection,
5 and is the clone identified as Nucleotide Sequence MV13-B1, proximal end. The human homolog is the (+) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

SEQ ID NO: 127 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

10 SEQ ID NO: 128 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

15 SEQ ID NO: 129 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E2, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

SEQ ID NO: 130 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E3, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

20 SEQ ID NO: 131 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E4, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

SEQ ID NO: 132 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E5, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

25 SEQ ID NO: 133 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E6, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

SEQ ID NO: 134 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E7, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

30 SEQ ID NO: 135 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E8, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

SEQ ID NO: 136 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E9, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

35 SEQ ID NO: 137 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E10, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC112218.2, transcription factor SMIF (HSA275986).

SEQ ID NO: 138 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E2, distal end. The human homolog is the (+) strand of GenBank Accession No. AL031293.1, KIAA1026.

5 SEQ ID NO: 139 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E3, distal end. The human homolog is the (+) strand of GenBank Accession No. AL035587.5, trinucleotide repeat containing 5 (TNRC5).

SEQ ID NO: 140 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E4, distal end. The human homolog is the (-) strand of GenBank Accession No. AC126182.2, homogentisate 1,2-dioxygenase (HGD).

10 SEQ ID NO: 141 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E5, distal end. The human homolog is the (+) strand of GenBank Accession No. AL591643.4, a region of chromosome Xq23-24.

15 SEQ ID NO: 142 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E6, distal end. The human homolog is the (-) strand of GenBank Accession No. AC113603.3, a region of chromosome 4p15.3.

SEQ ID NO: 143 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E7, distal end. The human homolog is the (+) strand of GenBank Accession No. AC011995.8, similar to LWamide neuropeptide precursor protein [Hydractinia echinata] (LOC129883).

20 SEQ ID NO: 144 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E8, distal end. The human homolog is the (-) strand of GenBank Accession No. AC084208.5, a region of chromosome 2q21.

25 SEQ ID NO: 145 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E9, distal end. The human homolog is the (-) strand of GenBank Accession No. AL391259.15, a region of chromosome Xp11.4, including ubiquitin specific protease 9 (USP9X).

SEQ ID NO: 146 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV13-E10, distal end. The human homolog is the (+) strand of GenBank Accession No. AC006397.1, LOC221829.

30 SEQ ID NO: 147 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-B2, distal end. The human homolog is the (+) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

35 SEQ ID NO: 148 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-B2, proximal end. The human homolog is the (+) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

SEQ ID NO: 149 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

SEQ ID NO: 150 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E2, proximal end. The human homolog is the (-) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

5 SEQ ID NO: 151 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E2, distal end. The human homolog is the (-) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

SEQ ID NO: 152 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E5, proximal end. The human homolog is the (-) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

10 SEQ ID NO: 153 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV8-E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

15 SEQ ID NO: 154 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV8-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. X14945.1, U3 small nuclear RNA gene.

SEQ ID NO: 155 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-B3, distal end. The human homolog is the (+) strand of GenBank Accession No. AL365203.19, integrin, beta 1 (ITGB1).

20 SEQ ID NO: 156 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-B3, proximal end. The human homolog is the (-) strand of GenBank Accession No. AL365203.19, ITGB1.

SEQ ID NO: 157 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E3, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL365203.19, ITGB1.

25 SEQ ID NO: 158 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E3, distal end. The human homolog is the (-) strand of GenBank Accession No. AL365203.19, ITGB1.

30 SEQ ID NO: 159 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AP001132.4, acrosomal vesicle protein 1 (ACRV1) and CHK1 checkpoint homolog (CHEK1).

SEQ ID NO: 160 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV7-E5, distal end. The human homolog is the (-) strand of GenBank Accession No. AK025453.1, prospero-related homeobox 1 (PROX1).

35 SEQ ID NO: 161 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E1, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

SEQ ID NO: 162 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E2, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

5 SEQ ID NO: 163 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E3, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

SEQ ID NO: 164 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E4, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

10 SEQ ID NO: 165 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E4, distal end. The human homolog is the (-) strand of GenBank Accession No. AL590543.8, between hypothetical proteins FLJ20627 and FLJ12910.

15 SEQ ID NO: 166 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E5, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

SEQ ID NO: 167 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E8, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

20 SEQ ID NO: 168 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E9, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

25 SEQ ID NO: 169 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E9, distal end. The human homolog is the (-) strand of GenBank Accession No. AL590543.8, between hypothetical proteins FLJ20627 and FLJ12910.

SEQ ID NO: 170 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E10, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

30 SEQ ID NO: 171 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E10, distal end. The human homolog is the (-) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

35 SEQ ID NO: 172 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV19-E2, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ20627.

SEQ ID NO: 173 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV19-E2, distal end. The human homolog is the (-) strand of GenBank Accession No. AL590543.8, between hypothetical proteins FLJ20627 and FLJ12910.

SEQ ID NO: 174 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AC105001.3, between PIN2-interacting protein 1 (PINX1) and SRY (sex-determining region Y)-box7 (SOX7).

5 SEQ ID NO: 175 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E2, distal end. The human homolog is the (-) strand of GenBank Accession No. AC009520.16, LOC131920.

10 SEQ ID NO: 176 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E3, distal end. The human homolog is the (-) strand of GenBank Accession No. AL596329.5, a region of chromosome 13q14.

SEQ ID NO: 177 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E5, distal end. The human homolog is the (+) strand of GenBank Accession No. AC023844.6, neurotrophic tyrosine kinase, receptor, type 3 (NTRK3).

15 SEQ ID NO: 178 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E7, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC024940.39, between TERA protein (TERA) and hypothetical protein FLJ13224.

20 SEQ ID NO: 179 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E7, distal end. The human homolog is the (+) strand of GenBank Accession No. AC024940.39, flanking TERA protein (TERA) and hypothetical protein FLJ13224.

25 SEQ ID NO: 180 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E8, distal end. The human homolog is the (-) strand of GenBank Accession No. AC084335.6, hypothetical gene LOC284260.

SEQ ID NO: 181 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E11, proximal end. The human homolog is the (+) strand of GenBank Accession No. AC073108.9, POM (POM121 homolog) and ZP3 fusion (POMZP3).

30 SEQ ID NO: 182 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV14-E11, distal end. The human homolog is the (-) strand of GenBank Accession No. AC073108.9, POM (POM121 homolog) and ZP3 fusion (POMZP3).

35 SEQ ID NO: 183 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV19-E4, distal end. The human homolog is the (+) strand of GenBank Accession No. AC087650.12, between DEAD/H box polypeptide 8 (DDX8) and similar to ribosomal protein L29 (cell surface heparin binding protein HIP) (LOC284064).

SEQ ID NO: 184 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-E2, distal end. The human homolog is the

(-) strand of GenBank Accession No. AC105285.3, LOC345307 and UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GALNT7).

SEQ ID NO: 185 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-E2, proximal end. The human homolog is
5 the (+) strand of GenBank Accession No. AC105285.3, LOC345307 and UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GALNT7).

SEQ ID NO: 186 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-B1, distal end. The human homolog is the
10 (+) strand of GenBank Accession No. AC105285.3, LOC345307 and UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GALNT7).

SEQ ID NO: 187 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-E3, distal end. The murine homolog is the
15 (+) strand of GenBank Accession No. NG_001440.1, *Mus musculus* 5S rRNA pseudogene (Rn5s-ps1).

SEQ ID NO: 188 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-E5, proximal end. The human homolog is
20 the (+) strand of GenBank Accession No. AL121886.22, between ribosomal protein L27a pseudogene (RPL27AP) and v-myb myeloblastosis viral oncogene homolog-like 2 (MYBL2).

SEQ ID NO: 189 is a nucleic acid sequence associated with viral, such as Ebola, infection,
25 and is the clone identified as Nucleotide Sequence MV28-E2, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL121886.22, between RPL27AP and MYBL2.

SEQ ID NO: 190 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E6, proximal end. The human homolog is the
30 (+) strand of GenBank Accession No. AL121886.22, between RPL27AP and MYBL2.

SEQ ID NO: 191 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E9, proximal end. The human homolog is the
35 (-) strand of GenBank Accession No. AL121886.22, between RPL27AP and MYBL2.

SEQ ID NO: 193 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-E6, distal end. The human homolog is the
30 (+) strand of GenBank Accession No. AP000711.4, Down's syndrome cell adhesion molecule like 1 (DSCAML1).

SEQ ID NO: 194 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-E7, distal end. The human homolog is the
35 (+) strand of GenBank Accession No. AL391555.19, LOC148529.

SEQ ID NO: 195 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV20-B4, distal end. The human homolog is the

(-) strand of GenBank Accession No. AC112129.4, Huntington-associated protein interacting protein (HAPIP).

SEQ ID NO: 196 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E1, proximal end. The human homolog is
5 the (-) strand of GenBank Accession No. Z69732.1, between LOC158525 and similar to RIKEN cDNA 1210001E11 (LOC347366).

SEQ ID NO: 197 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E3, proximal end. The human homolog is
10 the (-) strand of GenBank Accession No. Z69732.1, between LOC158525 and similar to RIKEN cDNA 1210001E11 (LOC347366).

SEQ ID NO: 198 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E5, proximal end. The human homolog is the (-) strand of GenBank Accession No. Z69732.1, between LOC158525 and similar to RIKEN cDNA 1210001E11 (LOC347366).

15 SEQ ID NO: 199 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E5, proximal end. The human homolog is the (-) strand of GenBank Accession No. Z69732.1, between LOC158525 and similar to RIKEN cDNA 1210001E11 (LOC347366).

20 SEQ ID NO: 200 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E8, proximal end. The human homolog is the (-) strand of GenBank Accession No. Z69732.1, between LOC158525 and similar to RIKEN cDNA 1210001E11 (LOC347366).

25 SEQ ID NO: 201 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E2, distal end. The human homolog is the (-) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ12910.

SEQ ID NO: 202 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E2, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ12910.

30 SEQ ID NO: 203 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E6, distal end. The human homolog is the (-) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ12910.

SEQ ID NO: 204 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E6, proximal end. The human homolog is the (+) strand of GenBank Accession No. AL590543.8, hypothetical protein FLJ12910.

35 SEQ ID NO: 205 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E7, distal end. The human homolog is the (+) strand of GenBank Accession No. AC005284.1, LOC350411.

SEQ ID NO: 206 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV22-E9, proximal end. The human homolog is

the (+) strand of GenBank Accession No. AP000505.1, between allograft inflammatory factor 1 (AIF1) and HLA-B associated transcript 2 (BAT2).

SEQ ID NO: 207 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV27-E1, distal end. The human homolog is the (-) strand of GenBank Accession No. AC008755.8, C19orf7.

SEQ ID NO: 208 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV27-E2, distal end. The human homolog is the (+) strand of GenBank Accession No. AC058791.4, between LOC346658 and LOC340349.

SEQ ID NO: 209 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV27-E2, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC058791.4, between LOC346658 and LOC340349.

SEQ ID NO: 210 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV27-E3, distal end. The human homolog is the (+) strand of GenBank Accession No. AC079030.13, a region of chromosome 12q21.

SEQ ID NO: 211 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV27-E3, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC139138.2, between LOC339248 and hypothetical protein FLJ22659.

SEQ ID NO: 212 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV27-E4, distal end. The human homolog is the (-) strand of GenBank Accession No. AL513550.9, between SR rich protein DKFZp564B0769 and hypothetical protein MGC14793.

SEQ ID NO: 213 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-B1, distal end. The human homolog is the (-) strand of GenBank Accession No. AP001160.4, hypothetical protein FLJ10439.

SEQ ID NO: 214 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-B1, proximal end. The human homolog is the (+) strand of GenBank Accession No. AP001160.4, hypothetical protein FLJ10439.

SEQ ID NO: 215 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-B3, distal end. The human homolog is the (+) strand of GenBank Accession No. AC090826.15, between cytochrome P450, family 11, subfamily A, polypeptide 1 (CYP11A1) and sema domain, immunoglobulin domain (Ig) and GPI membrane anchor, (semaphoring) 7A.

SEQ ID NO: 216 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-B3, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC090826.15, between cytochrome P450, family 11, subfamily A, polypeptide 1 (CYP11A1) and sema domain, immunoglobulin domain (Ig) and GPI membrane anchor, (semaphoring) 7A.

SEQ ID NO: 217 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E11, proximal end. The human homolog is the (+) strand of GenBank Accession No. AC090826.15, between cytochrome P450, family 11, subfamily A, polypeptide 1 (CYP11A1) and sema domain, immunoglobulin domain (Ig) and GPI membrane anchor, (semaphoring) 7A.

SEQ ID NO: 218 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E11, distal end. The human homolog is the (-) strand of GenBank Accession No. AC090826.15, between cytochrome P450, family 11, subfamily A, polypeptide 1 (CYP11A1) and sema domain, immunoglobulin domain (Ig) and GPI membrane anchor, (semaphoring) 7A.

SEQ ID NO: 219 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E1, proximal end. The human homolog is the (-) strand of GenBank Accession No. AC011500.7, ribosomal protein S16 (RPS16).

SEQ ID NO: 220 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E1, distal end. The human homolog is the (+) strand of GenBank Accession No. AC011500.7, ribosomal protein S16 (RPS16).

SEQ ID NO: 221 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E4, distal end. The human homolog is the (-) strand of GenBank Accession No. AC091172.11, between hypothetical protein DKFZp434H0115 and ATP citrate lyase (ACLY).

SEQ ID NO: 222 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E4, proximal end. The human homolog is the (+) strand of GenBank Accession No. AC091172.11, between hypothetical protein DKFZp434H0115 and ACLY.

SEQ ID NO: 223 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E7, distal end. The human homolog is the (+) strand of GenBank Accession No. AL035594.7, protein tyrosine phosphatase, receptor type, K (PTPRK).

SEQ ID NO: 224 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E7, proximal end. The human homolog is the (+) strand of GenBank Accession No. AC124857.2, calnexin (CANX) and (-) strand of GenBank Accession No. AL035594.7, protein tyrosine phosphatase, receptor type, K (PTPRK).

SEQ ID NO: 225 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E8, distal end. The human homolog is the (+) strand of GenBank Accession No. AC009144.5, cyclin M2 (CNNM2).

SEQ ID NO: 226 is a nucleic acid sequence associated with viral, such as Ebola, infection, and is the clone identified as Nucleotide Sequence MV28-E8, proximal end. The human homolog is the (+) strand of GenBank Accession No. AC011510.7, AXL receptor tyrosine kinase (AXL).

SEQ ID NO: 227 is a nucleic acid sequence showing GenBank Accession No. BC008947, *Homo sapiens* chromosome 10 open reading frame 3, mRNA (cDNA clone MGC:3422 IMAGE:3028566). This sequence is associated with viral infection, such as Ebola infection.

SEQ ID NO: 228 is an amino acid sequence encoded by SEQ ID NO: 227.

5 SEQ ID NO: 229 is a nucleic acid sequence showing GenBank Accession No. NM_018131, *Homo sapiens* chromosome 10 open reading frame 3 (C10orf3). This sequence is associated with viral infection, such as Ebola infection.

SEQ ID NO: 230 is an amino acid sequence encoded by SEQ ID NO: 229.

10 SEQ ID NO: 231 is a nucleic acid sequence showing GenBank Accession No. NM_013451, *Homo sapiens* fer-1-like 3, myoferlin (*C. elegans*) (FER1L3), transcript variant 1, mRNA. This sequence is associated with viral infection, such as Ebola infection.

SEQ ID NO: 232 is an amino acid sequence encoded by SEQ ID NO: 231.

SEQ ID NOS: 233 and 234 are exemplary complementary primers.

15 SEQ ID NOS: 235-237 are primer sequences used to sequence the shuttle clones as described in Example 2.

SEQ ID NOS: 238-241 are Rab9 siRNA sequences.

SEQ ID NOS: 242-245 are AXL receptor tyrosine kinase siRNA sequences.

SEQ ID NOS: 246-295 are beta-chimerin receptor tyrosine kinase RNAi sequences.

SEQ ID NOS: 296-345 are retinoblastoma binding protein 1 RNAi sequences.

20 SEQ ID NOS: 346-395 are *Homo sapiens* chromosome 10 open reading frame 3 RNAi sequences.

SEQ ID NOS: 396-445 are *Homo sapiens* fer-1-like 3, myoferlin (*C. elegans*), transcript variant 1 RNAi sequences.

25 SEQ ID NOS: 446-495 are *Homo sapiens* chromosome 10 open reading frame 3 (C10orf3) RNAi sequences.

SEQ ID NOS: 496-545 are malic enzyme RNAi sequences.

SEQ ID NOS: 546-595 are cadherin related 23 RNAi sequences.

SEQ ID NOS: 596-645 are sideroflexin 5 RNAi sequences.

SEQ ID NOS: 646-695 are polybromo 1 (PB1) RNAi sequences.

30 SEQ ID NOS: 696-720 are elongation factor for selenoprotein translation RNAi sequences.

SEQ ID NOS: 721-745 are integrin, beta 1 RNAi sequences.

SEQ ID NOS: 746-795 are huntingtin interacting protein 1 RNAi sequences.

SEQ ID NOS: 796-845 are cyclin M2 RNAi sequences.

35 DETAILED DESCRIPTION OF SEVERAL EMBODIMENTS

Abbreviations and Terms

The following explanations of terms and methods are provided to better describe the present disclosure and to guide those of ordinary skill in the art in the practice of the present disclosure. The singular forms "a," "an," and "the" refer to one or more than one, unless the context clearly dictates

otherwise. For example, the term "comprising a nucleic acid" includes single or plural nucleic acids and is considered equivalent to the phrase "comprising at least one nucleic acid." The term "or" refers to a single element of stated alternative elements or a combination of two or more elements, unless the context clearly indicates otherwise. For example, the phrase "a first nucleic acid or a second nucleic acid" refers to the first nucleic acid, the second nucleic acid, or a combination of both the first and second nucleic acids. As used herein, "comprises" means "includes." Thus, "comprising a promoter and an open reading frame," means "including a promoter and an open reading frame," without excluding additional elements.

Unless explained otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. The materials, methods, and examples are illustrative only and not intended to be limiting.

15	A = adenine
	C = cytosine
	DNA = deoxyribonucleic acid
	ds = double-stranded (for example, dsDNA)
	G = guanine
20	mg = milligram
	ng = nanogram
	PCR = polymerase chain reaction
	Pu = purine
	Py = pyrimidine
25	RNA = ribonucleic acid
	mRNA = messenger RNA
	MOI = multiplicity of infection
	siRNA = short interfering or interrupting RNA
	ss = single-stranded (for example, ssDNA)
30	T = thymine
	T _m = melting temperature
	U = uracil
	μg = microgram
	μl = microliter

35

Amplification of a nucleic acid. To increase the number of copies of a nucleic acid. Several methods can be used to amplify a nucleic acid, such as polymerase chain reaction (PCR). Other examples of amplification include, but are not limited to, strand displacement amplification (U.S. Patent No: 5,744,311); transcription-free isothermal amplification (U.S. Patent No: 6,033,881);

repair chain reaction amplification (WO 90/01069); ligase chain reaction amplification (European Patent Appl. 320 308); gap filling ligase chain reaction amplification (U.S. Patent No: 5,427,930); and NASBA™ RNA transcription-free amplification (U.S. Patent No: 6,025,134).

The amplification products ("amplicons") can be further processed, manipulated, or
5 characterized by electrophoresis, restriction endonuclease digestion, hybridization, nucleic acid sequencing, ligation, or other molecular biology techniques. Standard protocols can be modified. For example, PCR can be modified by using reverse transcriptase PCR (RT-PCR) to amplify RNA molecules.

10 **Antisense, Sense, and Antigene.** Antisense molecules are molecules that are specifically hybridizable or specifically complementary to either RNA or the plus strand of DNA. Sense molecules are molecules that are specifically hybridizable or specifically complementary to the minus strand of DNA. Antigene molecules are either antisense or sense molecules directed to a particular dsDNA target. These molecules can be used to interfere with gene expression.

15 Double-stranded DNA (dsDNA) has two strands, a 5' to 3' strand, referred to as the plus (+) strand, and a 3' to 5' strand (the reverse complement), referred to as the minus (-) strand. Because RNA polymerase adds nucleic acids in a 5' to 3' direction, the minus strand of the DNA serves as the template for the RNA during transcription. Thus, the RNA formed will have a sequence complementary to the minus strand and virtually identical to the plus strand, except that U is substituted for T in RNA molecules.

20 **Array.** An arrangement of biological samples or molecules, such as an arrangement of tissues, cells, or biological macromolecules (including, but not limited to, peptides or nucleic acids) in addressable locations on or in a substrate. The arrangement of molecules within the array can be regular, such as being arranged in uniform rows and columns, or irregular. The number of addressable locations within the array can vary, for example from a few (such as two or three) to
25 more than 50, 100, 200, 500, 1000, 10,000, or more. In certain examples, the array includes one or more molecules or samples occurring on the array a plurality of times (twice or more) to provide an added feature to the array, such as redundant activity or to provide internal controls. A "microarray" is an array that is miniaturized and evaluated or analyzed using microscopy.

30 Within an array, each arrayed sample or molecule is addressable, such that its location can be reliably and consistently determined within the at least two dimensions of the array. The location or address of each sample or molecule can be assigned when it is applied to the array, and a key or guide can be provided in order to correlate each location with the appropriate target sample or molecule position. Ordered arrays can be arranged in a symmetrical grid pattern or other patterns, for example, in radially distributed lines, spiral lines, or ordered clusters. Addressable arrays can be
35 computer readable; a computer can be programmed to correlate a particular address on the array with information about the sample at that position, such as hybridization or binding data, including signal intensity. In some exemplary computer readable formats, the individual samples or molecules in the array are arranged regularly (for example, in a Cartesian grid pattern), which can be correlated to address information by a computer.

The sample or molecule addresses on an array can assume many different shapes. For example, substantially square regions can be used as addresses within arrays, but addresses can be differently shaped, for example, substantially rectangular, triangular, oval, irregular, or another shape. The term "spot" refers generally to a localized placement of molecules, tissue or cells, and is not limited to a round or substantially round region or address.

Examples of macroarrays include the HistoTM-array and INSTA-blotTM lines of products available from Imgenix, Inc. (San Diego, CA) and the Max ArrayTM line of products available from Zymed Laboratories, Inc. (South San Francisco, CA), while exemplary microarrays include the various GeneChip[®] technologies and products available from Affymetrix, Inc. (Santa Clara, CA) and the HilightTM, Label StarTM, and Array-Ready Oligo Set lines of products available from Qiagen, Inc. (Valencia, CA).

β-chimerin. The term β-chimerin includes any β-chimerin gene, cDNA, RNA, or protein from any organism and is a β-chimerin that can function as a type of rho-GTPase. In some examples, β-chimerin is involved in viral infection.

Rho-GTPases are a family of small GTPases implicated as components of cellular signal transduction cascades. Signals that pass through rho-GTPase cascades can be initiated by the activation of cell surface proteins, such as growth factors. Functions of signaling cascades mediated by rho-GTPases, include, but are not limited to, alterations in cellular morphology which are linked to processes such as immune cell function, oncogenesis, metastasis and certain diseases (Peck, *FEBS Lett.* 528:27, 2002).

Examples of native β-chimerin nucleic acid sequences include, but are not limited to those shown in SEQ ID NOS: 21-22 (such as a target sequence associated with SEQ ID NOS: 21-22), as well as the protein sequence encoded thereby. This cell line remains CD4⁺ after exposure to HIV 1 and HIV 2 and is resistant to HIV infection. β-chimerin also includes variants, fusions, and fragments of the disclosed nucleic acid and amino acid sequences that retain β-chimerin biological activity.

Examples of β-chimerin amino acid sequences include, but are not limited to: Genbank Accession Nos: NM_004067 (mRNA) and NP_004058.1 (protein). In one example, a β-chimerin sequence includes a full-length wild-type (or native) sequence, as well as β-chimerin allelic variants, variants, fragments, homologs or fusion sequences that retain the ability to function as a type of rho-GTPase. In certain examples, β-chimerin has at least 80% sequence identity, for example at least 85%, 90%, 95%, or 98% sequence identity to a native β-chimerin.

cDNA (complementary DNA). A piece of DNA lacking internal, non-coding segments (introns) and transcriptional regulatory sequences. A cDNA also can contain untranslated regions (UTRs) that are responsible for translational control in the corresponding RNA molecule. cDNA can be produced using various methods, such as synthesis in the laboratory by reverse transcription from messenger RNA extracted from cells.

Complementary. Complementary binding occurs when the base of one nucleic acid molecule forms a hydrogen bond the base of another nucleic acid molecule. Normally, the base

adenine (A) is complementary to thymidine (T) and uracil (U), while cytosine (C) is complementary to guanine (G). For example, the sequence 5'-ATCG-3' of one ssDNA molecule can bond to 3'-TAGC-5' of another ssDNA to form a dsDNA.

Nucleic acid molecules can be complementary to each other even without complete
5 hydrogen-bonding of all bases of each molecule. By way of example only (and without limitation),
the ssDNA: 5'-GCTTGCCAAACCTACA-3' (SEQ ID NO: 233) is considered complementary to the
ssDNA 3'-CGAACGGCTGGATGT-5' (SEQ ID NO: 234) even though there is a mismatched base
pair (A-C rather than A-T or G-C) at the ninth position.

Conservative substitution: A substitution of an amino acid residue for another amino acid
10 residue having similar biochemical properties. Typically, conservative substitutions have little to no
impact on the biological activity of a resulting polypeptide. In a particular example, a conservative
substitution is an amino acid substitution in a peptide that does not substantially affect the biological
function of the peptide. A peptide can include one or more amino acid substitutions, for example 2-
10 conservative substitutions, 2-5 conservative substitutions, 4-9 conservative substitutions, such as
15 2, 5 or 10 conservative substitutions.

For example, a conservative substitution in a β -chimerin peptide (such as a peptide encoded
by a target sequence associated with SEQ ID NO: 21 or 22) does not substantially affect the ability of
 β -chimerin to confer resistance to HIV infection. In another example, a conservative substitution in a
Rab9 peptide (such as a peptide encoded by a target sequence associated with SEQ ID NOS: 118 or
20 119) is one that does not substantially affect the ability of Rab9 to confer resistance to infection by a
pathogen that can hijack a lipid raft, such as HIV or Ebola.

A polypeptide can be produced to contain one or more conservative substitutions by
manipulating the nucleotide sequence that encodes that polypeptide using, for example, standard
procedures such as site-directed mutagenesis or PCR. Alternatively, a polypeptide can be produced
25 to contain one or more conservative substitutions by using standard peptide synthesis methods. An
alanine scan can be used to identify which amino acid residues in a protein can tolerate an amino acid
substitution. In one example, the biological activity of the protein is not decreased by more than
25%, for example not more than 20%, for example not more than 10%, when an alanine, or other
conservative amino acid (such as those listed below), is substituted for one or more native amino
30 acids.

Examples of amino acids which can be substituted for an original amino acid in a protein
and which are regarded as conservative substitutions include, but are not limited to: Ser for Ala; Lys
for Arg; Gln or His for Asn; Glu for Asp; Ser for Cys; Asn for Gln; Asp for Glu; Pro for Gly; Asn or
Gln for His; Leu or Val for Ile; Ile or Val for Leu; Arg or Gln for Lys; Leu or Ile for Met; Met, Leu or
35 Tyr for Phe; Thr for Ser; Ser for Thr; Tyr for Trp; Trp or Phe for Tyr; and Ile or Leu for Val.

Further information about conservative substitutions can be found in, among other locations
in, Ben-Bassat *et al.*, (*J. Bacteriol.* 169:751-7, 1987), O'Regan *et al.*, (*Gene* 77:237-51, 1989), Sahin-

Toth *et al.*, (*Protein Sci.* 3:240-7, 1994), Hochuli *et al.*, (*Bio/Technology* 6:1321-5, 1988) and in standard textbooks of genetics and molecular biology.

5 **Ebola virus.** A highly contagious hemorrhagic virus named after a river in the Democratic Republic of the Congo (formerly Zaire) in Africa, where it was first recognized. Ebola is one of two members of a family of RNA viruses called the Filoviridae. There are four identified subtypes of Ebola virus. Three of the four have caused disease in humans: Ebola-Zaire, Ebola-Sudan, and Ebola-Ivory Coast. The fourth, Ebola-Reston, has caused disease in nonhuman primates, but not in humans.

10 **Ebola hemorrhagic fever (Ebola HF)** is a severe, often fatal disease in humans and nonhuman primates (for example, monkeys, gorillas, and chimpanzees) that is caused by Ebola virus infection. Diagnosing Ebola HF in a recently infected individual can be difficult because early symptoms, such as red eyes and a skin rash, are nonspecific to the virus and are seen in other subjects with diseases that occur much more frequently. Antigen-capture enzyme-linked immunosorbent assay (ELISA) testing, IgM ELISA, PCR, and virus isolation can be used to diagnose a case of Ebola HF within a few days after the onset of symptoms. Subjects tested later in the course of the disease, **15** or after recovery, can be tested for IgM and IgG antibodies. The disease also can be diagnosed retrospectively in deceased patients by using immunohistochemistry testing, virus isolation, or PCR.

20 **Encodes:** Unless evident from its context, includes DNA sequences that encode a polypeptide, as the term is typically used, as well as DNA sequences that are transcribed into inhibitory antisense molecules.

Expression: With respect to a gene sequence, refers to transcription of the gene and, as appropriate, translation of the resulting mRNA transcript to a protein. Thus, expression of a protein coding sequence results from transcription and translation of the coding sequence.

25 **Functional deletion:** A mutation, partial or complete deletion, insertion, or other variation made to a gene sequence that inhibits production of the gene product or renders the gene product non-functional. For example, a functional deletion of a Rab9 gene in a cell results in a cells having non-functional Rab9 protein, which results in the cell having an increase resistance to infection by a pathogen that uses a lipid raft.

30 **Gene.** A nucleic acid sequence that encodes a polypeptide under the control of a regulatory sequence, such as a promoter or operator. A gene includes an open reading frame encoding a polypeptide of the present disclosure, as well as exon and (optionally) intron sequences. An intron is a DNA sequence present in a given gene that is not translated into protein and is generally found between exons. The coding sequence of the gene is the portion transcribed and translated into a polypeptide (*in vivo*, *in vitro* or *in situ*) when placed under the control of an appropriate regulatory sequence. The boundaries of the coding sequence can be determined by a start codon at the 5' **35** (amino) terminus and a stop codon at the 3' (carboxyl) terminus. If the coding sequence is intended to be expressed in a eukaryotic cell, a polyadenylation signal and transcription termination sequence can be included 3' to the coding sequence.

Transcriptional and translational control sequences include, but are not limited to, DNA regulatory sequences such as promoters, enhancers, and terminators that provide for the expression of

the coding sequence, such as expression in a host cell. A polyadenylation signal is an exemplary eukaryotic control sequence. A promoter is a regulatory region capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding sequence. Additionally, a gene can include a signal sequence at the beginning of the coding sequence of a protein to be secreted or expressed on the surface of a cell. This sequence can encode a signal peptide, N-terminal to the mature polypeptide, which directs the host cell to translocate the polypeptide.

5 **Host Cell.** Any cell that can be infected with a virus or other pathogen, such as a bacterium. A host cell can be prokaryotic or eukaryotic, such as a cell from an insect, crustacean, mammal, bird, reptile, yeast, or a bacteria such as *E. coli*. Exemplary host cells include, but are not limited to, 10 mammalian B-lymphocyte cells. Examples of viruses include, but are not limited to HIV, influenza A, and Ebola.

15 The host cell can be part of an organism, or part of a cell culture, such as a culture of mammalian cells or a bacterial culture. A host nucleic acid is a nucleic acid present in a host cell that expresses a host protein. Decreasing or inhibiting the interaction between a host polypeptide or host nucleic acid and a virus or viral protein can occur *in vitro*, *in vivo*, and *in situ* environments.

20 **Human Immunodeficiency Virus (HIV).** A retrovirus that causes immunosuppression in humans and leads to a disease complex known as acquired immunodeficiency syndrome (AIDS). This immunosuppression results from a progressive depletion and functional impairment of T lymphocytes expressing the CD4 cell surface glycoprotein. The loss of CD4 helper/inducer T cell function may underlie the loss of cellular and humoral immunity leading to the opportunistic infections and malignancies seen in AIDS.

25 Depletion of CD4 T cells results from the ability of HIV to selectively infect, replicate in, and ultimately destroy these T cells (for example see Klatzmann *et al.*, *Science* 225:59, 1984). CD4 itself is an important component, and in some examples an essential component, of the cellular receptor for HIV.

30 HIV subtypes can be identified by particular number, such as HIV-1 and HIV-2. In the HIV life cycle, the virus enters a host cell in at least three stages: receptor docking, viral-cell membrane fusion, and particle uptake (D'Souza *et al.*, *JAMA* 284:215, 2000). Receptor docking begins with a gp120 component of a virion spike binding to the CD4 receptor on the host cell. Conformational changes in gp120 induced by gp120-CD4 interaction promote an interaction between gp120 and either CCR5 or CXCR4 cellular co-receptors. The gp41 protein then mediates fusion of the viral and target cell membranes. More detailed information about HIV can be found in Coffin *et al.*, *Retroviruses* (Cold Spring Harbor Laboratory Press, 1997).

35 **Hybridization.** Hybridization of a nucleic acid occurs when two complementary nucleic acid molecules undergo an amount of hydrogen bonding to each other. The stringency of hybridization can vary according to the environmental conditions surrounding the nucleic acids, the nature of the hybridization method, and the composition and length of the nucleic acids used. For example, temperature and ionic strength (such as Na⁺ concentration) can affect the stringency of hybridization. Calculations regarding hybridization conditions required for attaining particular

degrees of stringency are discussed in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); and Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes Part I*, Chapter 2 (Elsevier, New York, 1993).

5 The T_m is the temperature at which 50% of a given strand of nucleic acid is hybridized to its complementary strand. The T_m of a particular nucleic acid can be determined by various methods, such as observing the transition state between a single-stranded and double-stranded state during a temperature change, such as heating a dsDNA from about 30°C to about 100°C, and detecting when the dsDNA denatures to ssDNA. This can be accomplished by determining a melting profile for the
10 nucleic acid. For longer nucleic acid fragments, such as PCR products, the nearest-neighbor method can be used to determine T_m (Breslauer *et al.*, *Proc. Natl. Acad. Sci. USA* 83:3746-50, 1986). Additionally, MeltCalc software can be used to determine T_m (Schütz and von Ahsen, *Biotechniques* 30:8018-24, 1999).

15 For purposes of this disclosure, “stringent conditions” encompass conditions under which hybridization only will occur if there is less than 25% mismatch between the hybridization molecule and the target sequence. “Moderate stringency” conditions are those under which molecules with more than 25% sequence mismatch will not hybridize; conditions of “medium stringency” are those under which molecules with more than 15% mismatch will not hybridize, and conditions of “high stringency” are those under which sequences with more than 10% mismatch will not hybridize.
20 Conditions of “very high stringency” are those under which sequences with more than 5% mismatch will not hybridize.

25 Moderately stringent hybridization conditions are when the hybridization is performed at about 42°C in a hybridization solution containing 25 mM KPO₄ (pH 7.4), 5X SSC, 5X Denhart’s solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about 5x10⁷ cpm/µg), while the washes are performed at about 50°C with a wash solution containing 2X SSC and 0.1% sodium dodecyl sulfate.

30 Highly stringent hybridization conditions are when the hybridization is performed at about 42°C in a hybridization solution containing 25 mM KPO₄ (pH 7.4), 5X SSC, 5X Denhart’s solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about 5x10⁷ cpm/µg), while the washes are performed at about 65°C with a wash solution containing 0.2X SSC and 0.1% sodium dodecyl sulfate.

35 **Infection.** The entry, replication, insertion, lysis or other event or process involved with the pathogenesis of a virus or other infectious agent into a host cell. Thus, decreasing infection includes decreasing entry, replication, insertion, lysis, or other pathogenesis of a virus or other pathogen into a cell or subject, or combinations thereof. Infection includes the introduction of an infectious agent, such as a non-recombinant virus, recombinant virus, plasmid, bacteria, prion, eukaryotic microbe, or other agent capable of infecting a host, such as the cell of a subject.

 In another example, infection is the introduction of a recombinant vector into a host cell via transduction, transformation, transfection, or other method. Vectors include, but are not limited to,

viral, plasmid, cosmid, and artificial chromosome vectors. For example, a recombinant vector can include an antisense molecule, RNAi molecule, or siRNA that recognizes any target sequences associated with SEQ ID NOS: 1-227, 229, and 231, or variants, fusions, or fragments thereof, as well as SEQ ID NOS: 1-227, 229, and 231 themselves.

5 **Influenza virus.** A virus that causes respiratory disease or influenza ("the flu") and can lead to a secondary infection in the host, such as a bacterial infection of the lungs. Three types of influenza are currently known: influenza A, influenza B, and influenza C. Influenza A is the most common form of the virus and is capable of infection humans and non-human animals, such as pigs, horses, chickens, ducks and other birds.

10 The viral genome includes eight RNA molecules. HA, which encodes hemagglutinin (three hemagglutinin subtypes: H1, H2, and H3); M, which encodes two matrix proteins based on two different open reading frames within the nucleic acid sequence; NA encodes for neuraminidase; NP encodes the nucleoprotein; NS encodes two non-structural proteins based on different open reading frames within the nucleic acid sequence; and three genes that encode RNA polymerases (PA, PB1, 15 PB2). The influenza virus can be categorized into subtypes on the bases of the surface glycoproteins.

20 The replication cycle of the influenza virus begins with binding of the viral hemagglutinin molecules to the surface carbohydrate of epithelial cell of a host cell, which draws the virus into the cell by receptor-mediated endocytosis. The viral membrane fuses with the endocytotic vesicle membrane, allowing the RNA molecules of the viral genome to enter the interior of the cell where these molecules later enter the cell nucleus and are replicated into viral-complementary RNA and new viral RNA and transcribed into viral mRNA, which are transported into the cytosol where they are translated into the proteins of new viral particles. After viral particles are assembled into new viruses, the neuraminidase glycoproteins proteins aid in the budding of the viruses from the cellular membrane of the host cell, thus releasing new viruses capable of infecting other host cells.

25 **Isolated:** An "isolated" biological component (such as a nucleic acid or protein) has been substantially separated, produced apart from, or purified away from other biological components in the cell of the organism in which the component naturally occurs, such as other chromosomal and extrachromosomal DNA and RNA, and proteins. Nucleic acids and proteins which have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term 30 also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids, proteins and peptides.

35 **Nucleic acid.** A deoxyribonucleotide or ribonucleotide polymer in either single (ss) or double stranded (ds) form, and can include analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. In some examples, a nucleic acid is a nucleotide analog.

 Unless otherwise specified, any reference to a nucleic acid molecule includes the reverse complement of nucleic acid. Except where single-strandedness is required by the text herein (for example, a ssRNA molecule), any nucleic acid written to depict only a single strand encompasses both strands of a corresponding double-stranded nucleic acid. For example, depiction of a plus-strand

of a dsDNA also encompasses the complementary minus-strand of that dsDNA. Additionally, reference to the nucleic acid molecule that encodes a specific protein, or a fragment thereof, encompasses both the sense strand and its reverse complement.

In particular examples, a nucleic acid includes a nucleotide sequence shown in any of SEQ ID NOS: 1-227, 229, and 231, or a variant, fragment, or fusion thereof. In other examples, a nucleic acid has a nucleotide sequence including a target sequence associated with SEQ ID NOS: 1-227, 229, and 231, or a variant, fragment, or fusion thereof, such as the corresponding cDNA or mRNA of SEQ ID NOS: 1-227, 229, and 231.

The fragment can be any portion of the nucleic acid corresponding to at least 5 contiguous bases from any target nucleic acid sequence associated with SEQ ID NOS: 1-227, 229, and 231, for example at least 20 contiguous bases, at least 50 contiguous bases, at least 100 contiguous bases, at least 250 contiguous bases, or even at least 500 or more contiguous bases. A fragment can be chosen from a particular portion of any of the target sequences associated with SEQ ID NOS: 1-227, 229, and 231, such as a particular half, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, or smaller portion of any of the target sequences associated with SEQ ID NOS: 1-227, 229, and 231. Fragments of the nucleic acids described herein can be used as probes and primers.

Oligonucleotide. A linear polynucleotide (such as DNA or RNA) sequence of at least 9 nucleotides, for example at least 15, 18, 24, 25, 30, 50, 100, 200 or even 500 nucleotides long. In particular examples, an oligonucleotide is about 6-50 bases, for example about 10-25 bases, such as 12-20 bases.

An oligonucleotide analog refers to moieties that function similarly to oligonucleotides, but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules.

Open reading frame (ORF). A series of nucleotide triplets (codons) coding for amino acids without any internal termination codons. These sequences are usually translatable into a peptide.

Operably linked. A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in the same reading frame.

Pathogen: A disease-producing agent. Examples include, but are not limited to viruses, bacteria, and fungi.

Pharmaceutical agent or drug: A chemical compound or composition capable of inducing a desired therapeutic or prophylactic effect when administered to a subject, alone or in combination with another therapeutic agent(s) or pharmaceutically acceptable carriers. In a particular example, a

pharmaceutical agent decreases or even inhibits infection of a cell, such as the cell of a subject, by a pathogen, such as a virus.

5 **Polymorphism.** A polymorphism exists when two or more versions of a nucleic acid sequence exist within a population of subjects. For example, a polymorphic nucleic acid can be one where the most common allele has a frequency of 99% or less. Different alleles can be identified according to differences in nucleic acid sequences, and genetic variations occurring in more than 1% of a population (which is the commonly accepted frequency for defining polymorphism) are useful polymorphisms for certain applications.

10 The allelic frequency (the proportion of all allele nucleic acids within a population that are of a specified type) can be determined by directly counting or estimating the number and type of alleles within a population. Polymorphisms and methods of determining allelic frequencies are discussed in Hartl, D.L. and Clark, A.G., *Principles of Population Genetics*, Third Edition (Sinauer Associates, Inc., Sunderland Massachusetts, 1997), particularly in chapters 1 and 2.

15 **Preventing or treating a disease:** "Preventing" a disease refers to inhibiting the full development of a disease, for example preventing development of a viral infection. "Treatment" refers to a therapeutic intervention that ameliorates a sign or symptom of a disease or pathological condition related to a viral infection, such as inhibiting or decreasing viral infection.

20 **Probes and primers.** A probe includes an isolated nucleic acid attached to a detectable label or other reporter molecule. Typical labels include, but are not limited to radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed for example in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989) and Ausubel *et al.* (*In Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998).

25 Primers are short nucleic acid molecules, such as DNA oligonucleotides ten nucleotides or more in length. Longer DNA oligonucleotides can be about 15, 20, 25, 30 or 50 nucleotides or more in length. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then the primer extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a 30 nucleic acid sequence, for example by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods.

35 Nucleic acid probes and primers can be prepared based on the nucleic acid molecules of the target sequences associated with SEQ ID NOS: 1-227, 229, and 231, as indicators of resistance to infection. Probes and primers can be based on fragments or portions of these nucleic acid molecules, or on the reverse complement of these sequences, as well as probes and primers to 5' or 3' regions of the nucleic acids.

The specificity of a probe or primer increases with its length. Thus, for example, a primer that includes 30 consecutive nucleotides of a β -chimerin or Rab9 gene will anneal to a target sequence, such as another homolog of a β -chimerin or Rab9 gene, respectively, with a higher specificity than a

corresponding primer of only 15 nucleotides. Thus, to obtain greater specificity, probes and primers can be selected that include at least 20, 25, 30, 35, 40, 45, 50 or more consecutive nucleotides of a nucleic acid disclosed herein.

Protein coding sequence or a sequence that encodes a peptide: A nucleic acid sequence 5 that is transcribed (in the case of DNA) and is translated (in the case of mRNA) into a peptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and 10 even synthetic DNA sequences. A transcription termination sequence is usually located 3' to the coding sequence.

Purified. The term purified does not require absolute purity; rather, it is a relative term. Thus, for example, a purified peptide preparation is one in which the peptide or protein is more enriched than the peptide or protein is in its environment within a cell, such that the peptide is 15 substantially separated from cellular components (nucleic acids, lipids, carbohydrates, and other polypeptides) that may accompany it. In another example, a purified peptide preparation is one in which the peptide is substantially-free from contaminants, such as those that might be present following chemical synthesis of the peptide.

In one example, a peptide is purified when at least 60% by weight of a sample is composed 20 of the peptide, for example when 75%, 95%, or 99% or more of a sample is composed of the peptide, such as a β-chimerin or Rab9 peptide. Examples of methods that can be used to purify proteins, include, but are not limited to the methods disclosed in Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989, Ch. 17). Protein purity can be determined by, for example, polyacrylamide gel electrophoresis of a protein sample, followed by visualization of 25 a single polypeptide band upon staining the polyacrylamide gel; high-pressure liquid chromatography; sequencing; or other conventional methods.

Rab9: The term Rab9 includes any Rab9 gene, cDNA, RNA, or protein from any organism and that is a Rab9 that can transport late endosomes to trans-golgi and function as a ras-like GTPase. In some examples, Rab9 is involved in lipid raft formation.

30 Examples of native Rab9 nucleic acid sequences include, but are not limited to, target sequences associated with SEQ ID NOS: 118 and 119. Examples of Rab9 amino acid sequences include, but are not limited to: Genbank Accession Nos: BC017265.2 and NM_004251.3 (cDNA) as well as P51151 and AAH17265 (proteins). In one example, a Rab9 sequence includes a full-length wild-type (or native) sequence, as well as Rab9 allelic variants, variants, fragments, homologs or 35 fusion sequences that retain the ability to transport late endosomes to trans-golgi. In certain examples, Rab9 has at least 80% sequence identity, for example at least 85%, 90%, 95%, or 98% sequence identity to a native Rab9.

In other examples, Rab9 has a sequence that hybridizes to a sequence set forth in GenBank Accession No. BC017265.2 or NM_004251.3, and retains Rab9 activity.

Recombinant. A recombinant nucleic acid or protein is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished, for example, by chemical synthesis or by the artificial manipulation of isolated segments of nucleic acids or proteins,
5 for example, by genetic engineering techniques.

RNA interference (RNAi): A post-transcriptional gene silencing mechanism mediated by double-stranded RNA (dsRNA). Introduction of dsRNA into cells, such as RNAi compounds or siRNA compounds, induces targeted degradation of RNA molecules with homologous sequences.
10 RNAi compounds are typically longer than an siRNA molecule. For example, an RNAi molecule can be at least about 25 nucleic acids, at least about 27 nucleic acids, or even at least about 400 nucleotides in length.

RNAi compounds can be used to modulate transcription, for example, by silencing genes, such as Rab9, β -chimerin, or combinations thereof. In certain examples, an RNAi molecule is directed against a certain target gene, such as Rab9, β -chimerin, or combinations thereof, and is used
15 to decrease viral infection.

Sequence identity: The similarity between nucleic acid or amino acid sequences is expressed in terms of the similarity between the sequences. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are. Homologs or variants of a protein or nucleic acid disclosed herein,
20 such as target sequences associated with SEQ ID NOS: 1-232, and their corresponding cDNA and protein sequences, will possess a relatively high degree of sequence identity when aligned using standard methods.

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman, *Adv. Appl. Math.* 2:482, 25 1981; Needleman and Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85:2444, 1988; Higgins and Sharp, *Gene* 73:237-44, 1988; Higgins and Sharp, *CABIOS* 5:151-3, 1989; Corpet *et al.*, *Nucl. Acids Res.* 16:10881-90, 1988; Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85:2444, 1988; and Altschul *et al.*, *Nature Genet.* 6:119-29, 1994.

The NCBI Basic Local Alignment Search Tool (BLASTTM) (Altschul *et al.*, *J. Mol. Biol.* 30 215:403-10, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx.

Variants of a peptide, such as a peptide encoded by any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, as well as any target sequence associated with SEQ ID NOS: 228, 230, 35 and 232, are typically characterized by possession of at least 70% sequence identity counted over the full length alignment with the amino acid sequence encoded by any target sequence associated with SEQ ID NOS: 1-227, 229, or 231, using the NCBI Blast 2.0, gapped blastp set to default parameters. For comparisons of amino acid sequences of greater than about 30 amino acids, the Blast 2 sequences function is employed using the default BLOSUM62 matrix set to default parameters, (gap existence

cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment is performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the reference sequences will show increasing percentage identities when assessed 5 by this method, such as at least 80%, at least 90%, at least 95%, at least 98%, or even at least 99% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs and variants will typically possess at least 80% sequence identity over short windows of 10-20 amino acids, and may possess sequence identities of at least 85%, at least 90%, at least 95%, or at least 98% depending on their similarity to the reference sequence. Methods for determining 10 sequence identity over such short windows are described at the website that is maintained by the National Center for Biotechnology Information in Bethesda, Maryland. One of skill in the art will appreciate that these sequence identity ranges are provided for guidance only; it is entirely possible that strongly significant homologs could be obtained that fall outside of the ranges provided.

Similar methods can be used to determine the sequence identity between two or more 15 nucleic acids. To compare two nucleic acid sequences, the BLASTN options can be set as follows: -i is set to a file containing the first nucleic acid sequence to be compared (such as C:\seq1.txt); -j is set to a file containing the second nucleic acid sequence to be compared (such as C:\seq2.txt); -p is set to blastn; -o is set to any desired file name (such as C:\output.txt); -q is set to -1; -r is set to 2; and all other options are left at their default setting. For example, the following command can be used to 20 generate an output file containing a comparison between two sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt -q -1 -r 2.

Once aligned, the number of matches is determined by counting the number of positions where an identical nucleotide or amino acid residue is presented in both sequences. The percent sequence identity is determined by dividing the number of matches either by the length of the 25 sequence set forth in the identified sequence, or by an articulated length (for example, 100 consecutive nucleotides or amino acid residues from a sequence set forth in an identified sequence), followed by multiplying the resulting value by 100. For example, a nucleic acid sequence that has 1166 matches when aligned with a test sequence having 1154 nucleotides is 75.0 percent identical to the test sequence (for example, $1166 \div 1554 \times 100 = 75.0$). The percent sequence identity value is 30 rounded to the nearest tenth. For example, 75.11, 75.12, 75.13, and 75.14 are rounded down to 75.1, while 75.15, 75.16, 75.17, 75.18, and 75.19 are rounded up to 75.2. The length value will always be an integer. In another example, a target sequence containing a 20-nucleotide region that aligns with 20 consecutive nucleotides from an identified sequence as follows contains a region that shares 75 percent sequence identity to that identified sequence (for example, $15 \div 20 \times 100 = 75$).

35

1

20

Target Sequence:

AGGTCGTGTACTGTCAGTCA

| | | | | | | | | | | | | | | |

Identified Sequence:

ACGTGGTGAACTGCCAGTGA

The nucleic acids disclosed herein include nucleic acids have nucleotide sequences that are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to the nucleotide sequence of any target sequence associated with SEQ ID NOS: 1-227, 229, and 231. In particular examples, a nucleic acid is substantially similar to the nucleotide sequence of any target sequence associated with SEQ ID NOS: 1-227, 229, and 231. A first nucleic acid is "substantially similar" to a second nucleic acid if, when the first nucleic acid is optimally aligned (with appropriate nucleotide deletions or gap insertions) with the second nucleic acid (or its complementary strand) and there is nucleotide sequence identity of at least about 90%, for example at least about 95%, at least 98% or at least 99% identity. An alternative indication that two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid molecules that all encode substantially the same protein.

Short interfering or interrupting RNA (siRNA). Double-stranded RNAs that can induce sequence-specific post-transcriptional gene silencing, thereby decreasing or even inhibiting gene expression. In some examples, siRNA molecules are about 19-23 nucleotides in length, such as at least 21 nucleotides, for example at least 23 nucleotides.

In one example, siRNA triggers the specific degradation of homologous RNA molecules, such as mRNAs, within the region of sequence identity between both the siRNA and the target RNA. For example, WO 02/44321 discloses siRNAs capable of sequence-specific degradation of target mRNAs when base-paired with 3' overhanging ends. The direction of dsRNA processing determines whether a sense or an antisense target RNA can be cleaved by the produced siRNA endonuclease complex. Thus, siRNAs can be used to modulate transcription, for example, by silencing genes, such as Rab9, β -chimerin, or combinations thereof. The effects of siRNAs have been demonstrated in cells from a variety of organisms, including *Drosophila*, *C. elegans*, insects, frogs, plants, fungi, mice and humans (for example, WO 02/44321; Gitlin *et al.*, *Nature* 418:430-4, 2002; Caplen *et al.*, *Proc. Natl. Acad. Sci.* 98:9742-9747, 2001; and Elbashir *et al.*, *Nature* 411:494-8, 2001).

In certain examples, siRNAs are directed against certain target genes, such as Rab9, β -chimerin, or combinations thereof, to confirm results of the gene-trap method used against the same nucleic acid sequence.

Specific binding agent. An agent that binds substantially only to a defined target. For example, a protein-specific binding agent binds substantially only the specified protein and a nucleic acid specific binding agent binds substantially only the specified nucleic acid.

As used herein, the term "protein [X] specific binding agent" includes anti-[X] protein antibodies (including polyclonal or monoclonal antibodies and functional fragments thereof) and other agents (such as soluble receptors) that bind substantially only to the [X] protein. In this context, [X] refers to any specific or designated protein, for instance β -chimerin, Rab9, or any protein listed in Table

1 or encoded by a target sequence associated with SEQ ID NOS: 1-227, 229, and 231 (including variants, fragments, and fusions thereof).

Anti-[X] protein antibodies can be produced using standard procedures such as those described in Harlow and Lane (*Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press: 5 Cold Spring Harbor, 1998). Antibodies can be polyclonal or monoclonal antibodies, humanized antibodies, Fab fragments, F(ab')2 fragments, single chain antibodies, or chimeric antibodies. For example, polyclonal antibodies can be produced by immunizing a host animal by injection with polypeptides described herein, including the target sequences associated with SEQ ID NOS: 1-227, 229, 10 231 (or variants, fragments, or fusions thereof). The production of monoclonal antibodies can be accomplished by a variety of methods, such as the hybridoma technique (Kohler and Milstein, *Nature* 256:495-7, 1975), the human B-cell technique (Kosbor *et al.*, *Immunology Today* 4:72, 1983), or the EBV-hybridoma technique (Cole *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96, 1983). Additionally, chimeric antibodies can be produced (for example, see Morrison *et al.*, *J. Bacteriol.* 159:870, 1984; Neuberger *et al.*, *Nature* 312:604-8, 1984; and Takeda *et al.*, *Nature* 15 314:452-4, 1985), as well as single-chain antibodies (for example, see U.S. Pat. Nos: 5,476,786; 5,132,405; and 4,946,778).

The determination that a particular agent binds substantially only to the specified protein readily can be made by using or adapting routine procedures. For example, Western blotting can be used to determine that a given protein binding agent, such as an anti-[X] protein monoclonal antibody, 20 binds substantially only to the [X] protein. Other assays include, but are not limited to, competitive and non-competitive homogenous and heterogeneous enzyme-linked immunosorbent assays (ELISA) as symmetrical or asymmetrical direct or indirect detection formats; "sandwich" immunoassays; immunodiffusion assays; *in situ* immunoassays (for example, using colloidal gold, enzyme or radioisotope labels); agglutination assays; complement fixing assays; immunoelectrophoretic assays; 25 enzyme-linked immunospot assays (ELISPOT); radioallergosorbent tests (RAST); fluorescent tests, such as used in fluorescent microscopy and flow cytometry; Western, grid, dot or tissue blots; dip-stick assays; halogen assays; or antibody arrays (for example, see O'Meara and Tovey, *Clin. Rev. Allergy Immunol.*, 18:341-95, 2000; Sambrook *et al.*, 2001, Appendix 9; Simonnet and Guilloteau, in: *Methods of Immunological Analysis*, Masseyeff *et al.* (Eds.), VCH, New York, 1993, pp. 270-388).

30 A specific binding agent also can be labeled for direct detection (see Chapter 9, Harlow and Lane, *Antibodies: A Laboratory Manual*. 1988). Suitable labels include (but are not limited to) enzymes (such as alkaline phosphatase (AP) or horseradish peroxidase (HRP)), fluorescent labels, colorimetric labels, radioisotopes, chelating agents, dyes, colloidal gold, ligands (such as biotin), and chemiluminescent agents.

35 Shorter fragments of antibodies can also serve as specific binding agents. For instance, Fabs, Fvs, and single-chain Fvs (SCFvs) that bind to a specified protein would be specific binding agents. These antibody fragments include: (1) Fab, the fragment containing a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an

antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) (Fab')₂, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; (4) F(ab')₂, a dimer of two Fab' fragments held
5 together by two disulfide bonds; (5) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (6) single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Methods of making these fragments are routine. For
10 example, construction of Fab expression libraries permits the rapid and easy identification of monoclonal Fab fragments with the desired specificity for a protein described herein.

Subject: Living multi-cellular vertebrate organisms, including human and veterinary subjects, such as cows, pigs, horses, dogs, cats, birds, reptiles, and fish.

Target sequences associated with SEQ ID NO: When used herein, this phrase refers to
15 any nucleic acid sequence, amino acid sequence, or combination of nucleic acid and amino acid sequences, that are involved in viral infection, and therefore serve as targets for inhibiting viral infection, and which are or include a specified SEQ ID NO, are involved in the expression of the SEQ ID NO, or are peptide (including protein) sequences that are expressed by such specified SEQ ID NO. Although a target sequence may refer to a SEQ ID NO of a sequence obtained from a particular
20 species, the target sequences also include homologues of the sequence from other related species, such as other mammals. For example, the phrase "target sequences associated with SEQ ID NO. X" can refer to the entire gene sequence of which the particular SEQ ID NO X is a part, the appropriate coding sequence, a promoter sequence associated with the gene, or the corresponding protein sequence, as well as variants, fragments, homologues, and fusions thereof that retain the activity of
25 the native sequence.

For example, when using the phrase "sequences associated with SEQ ID NOS: 21-22," this term encompasses β-chimerin genomic sequences, endogenous promoter sequences that promote the expression of β-chimerin, coding sequences, and β-chimerin proteins, as well as variants, fragments homologues and fusions thereof that retain the activity of the native sequence. A particular cDNA sequence associated with SEQ ID NOS: 21-22 is provided in GenBank Accession No. NM_004067, and a particular protein sequence associated with SEQ ID NOS: 21-22 is provided in NP_004058.1.
30

The term "a GenBank Accession No. associated with SEQ ID NO. X" refers to a GenBank Accession No. that includes SEQ ID NO. X, or is a homolog of SEQ ID NO: X from another mammal, for example a human homolog. The GenBank Accession No. may, in some examples, also identify a coding sequence of an open reading frame, and the sequence of the protein encoded by SEQ ID NO. X.
35

Although sequences are provided herein that encode (or are included within sequences that encode) host proteins that are involved in viral infection, it should be understood that the ultimate goal is to interfere with the activity of the protein that has been identified to be involved in viral

pathogenesis. Such interference can be at either the level of the nucleic acid that encodes the protein (for example by reducing or otherwise disrupting expression of the protein), or at the level of the protein itself (for example by interfering with the activity of the protein, or its interaction with the virus). The disclosure of specific techniques for achieving these goals in particular species should not 5 be interpreted to limit the method to these particular techniques, or to particular species in which the viral interaction is first identified. The identification of the viral interaction in one species indicates the importance of the interaction between the virus and the protein in that species, as well as the interaction of the virus with homologues of that protein in other species.

Target sequence of a nucleic acid: A portion of a nucleic acid that, upon hybridization to a 10 therapeutically effective oligonucleotide or oligonucleotide analog, results in reduction or even inhibition of infection by an infectious agent. An antisense or a sense molecule can be used to target a portion of dsDNA, since either can interfere with the expression of that portion of the dsDNA. The antisense molecule can bind to the plus strand, and the sense molecule can bind to the minus strand. Thus, target sequences can be ssDNA, dsDNA, and RNA.

Therapeutically active molecule: An agent, such as a protein, antibody or nucleic acid, 15 that can decrease expression of a host protein involved in viral infection (such as those listed in Table 1 or target sequences associated with any of SEQ ID NOS: 1-232, or can decrease an interaction between a host protein involved in viral infection and a viral protein, such as HIV, Ebola, or influenza A, as measured by clinical response (for example, a decrease in infection by a virus, such as an 20 inhibition of infection). Therapeutically active agents also include organic or other chemical compounds that mimic the effects of the therapeutically effective peptide or nucleic acids.

Therapeutically Effective Amount: An amount of a pharmaceutical preparation that alone, or together with an additional therapeutic agent(s), induces the desired response. The preparations disclosed herein are administered in therapeutically effective amounts.

In one example, a desired response is to decrease or inhibit viral infection of a cell, such as a 25 cell of a subject. Viral infection does not need to be completely inhibited for the pharmaceutical preparation to be effective. For example, a pharmaceutical preparation can decrease viral infection by a desired amount, for example by at least 20%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, or even at least 100%, as compared to an amount of 30 viral infection in the absence of the pharmaceutical preparation. This decrease or inhibition can result in halting or slowing the progression of, or inducing a regression of a pathological condition caused by the viral infection, or which is capable of relieving signs or symptoms caused by the condition.

In another or additional example, it is an amount sufficient to partially or completely 35 alleviate symptoms of viral infection within a host subject. Treatment can involve only slowing the progression of the infection temporarily, but can also include halting or reversing the progression of the infection permanently.

Effective amounts of the therapeutic agents described herein can be determined in many different ways, such as assaying for a reduction in the rate of infection of cells or subjects, a reduction in the viral load within a host, improvement of physiological condition of an infected subject, or

increased resistance to infection following exposure to the virus. Effective amounts also can be determined through various *in vitro*, *in vivo* or *in situ* assays, including the assays described herein.

The disclosed therapeutic agents can be administered in a single dose, or in several doses, for example daily, during a course of treatment. However, the effective amount of can be dependent 5 on the source applied (for example a nucleic acid isolated from a cellular extract versus a chemically synthesized and purified nucleic acid), the subject being treated, the severity and type of the condition being treated, and the manner of administration. In addition, the disclosed therapeutic agents can be administered alone, or in the presence of a pharmaceutically acceptable carrier, or in the presence of other therapeutic agents, for example other anti-viral agents.

10 **Transduced and Transformed:** A virus or vector "transduces" or "transfects" a cell when it transfers nucleic acid into the cell. A cell is "transformed" by a nucleic acid transduced into the cell when the DNA becomes stably replicated by the cell, either by incorporation of the nucleic acid into the cellular genome, or by episomal replication. As used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, 15 including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

20 **Transfected:** A transfected cell is a cell into which has been introduced a nucleic acid molecule by molecular biology techniques. The term transfection encompasses all techniques by which a nucleic acid molecule can be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

25 **Transgene:** An exogenous nucleic acid sequence supplied by a vector. In one example, a transgene includes any target sequence associated with SEQ ID NOS: 1-227, 229, 231 (or variants, fragments, or fusions thereof), for example a nucleic acid that encodes a beta-chimerin or Rab9.

30 **Variants, fragments or fusions:** The disclosed nucleic acid sequences, such as target sequences associated with SEQ ID NOS: 1-227, 229, and 231, and the proteins encoded thereby, include variants, fragments, and fusions thereof that retain the native biological activity (such as playing a role in viral infection). DNA sequences which encode for a protein or fusion thereof, or a fragment or variant of thereof can be engineered to allow the protein to be expressed in eukaryotic cells or organisms, bacteria, insects, and/or plants. To obtain expression, the DNA sequence can be altered and operably linked to other regulatory sequences. The final product, which contains the regulatory sequences and the therapeutic protein, is referred to as a vector. This vector can be introduced into eukaryotic, bacteria, insect, and/or plant cells. Once inside the cell the vector allows the protein to be produced.

35 One of ordinary skill in the art will appreciate that the DNA can be altered in numerous ways without affecting the biological activity of the encoded protein. For example, PCR can be used to produce variations in the DNA sequence which encodes a protein. Such variants can be variants optimized for codon preference in a host cell used to express the protein, or other sequence changes that facilitate expression.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication, and can also include one or more selectable marker genes and other genetic elements. An insertional vector is capable of inserting itself into a host nucleic acid.

5 For example, recombinant lambda-phage vectors of host genomes (Coffin *et al.*, *Retroviruses*, Chapter 5).

Wild-type. A naturally occurring, non-mutated version of a nucleic acid sequence. Among multiple alleles, the allele with the greatest frequency within the population is usually (but not necessarily) the wild-type. The term "native" can be used as a synonym for "wild-type."

10

Nucleic Acids and Proteins Involved in Viral Infection

Examples of host nucleic acids and proteins that play a role in viral infection have been identified and are summarized in Table 1. These nucleic acids and proteins offer new targets for therapies that reduce or even inhibit or prevent viral infection, and offer new strategies for assessing 15 the risk of infection among certain populations. While the target genes were identified in an assay using the recited virus, it is appreciated that infections agents such as viruses will share common pathways. Thus, the host sequences set forth below can be interfered with to decrease infection in a host cell.

20 Examples of viruses that can be inhibited are described in Virology, Volumes 1 and 2 by Bernard Fields, Second Edition, 1990, Raven Press. Exemplary viruses include, but are not limited to members of the family: Picornaviridae (such as Poliovirus, Coxsackievirus, Echovirus, Rhinovirus, and Hepatitis A and E); Calciviridae (such as Norwalk and related viruses); Togaviridae and Flaviviridae (such as hepatitis C, Alphavirus, and Rubella); Coronaviridae (such as SARS); Rhabdoviridae (such as Rabies); Filoviridae (such as Marburg and Ebola); Paramyxoviridae (such as 25 Parainfluenza, Mumps, Measles, Hydra and Respiratory Syncytial virus); Orthomyxoviridae; Bunyaviridae (including all subtypes and strains); Arenaviridae (such as lymphocytic choriomeningitis virus and lassa fever and related viruses); Reoviridae (such as Reovirus and Rotavirus); Retroviridae (such as HTLV, HIV, and Lentivirus); Papoviridae (such as Polyoma and Papilloma); Adenoviridae (such as Adenovirus); Parvoviridae (such as Parvovirus); Herpesviridae 30 (such as Herpes 1 and 2, Cytomegalovirus, Varicella-Zoster, Kaposi sarcoma related virus (HHV9), Epstein Barr Virus, and HHV6-7 (roseolavirus)); Poxviridae (such as Pox); Hepadnaviridae (such as Hepatitis B); as well as Hepatitis D virus, Hanta virus, and newly identified infectious agents.

Table 1: Examples of Host Genes and Proteins Implicated in Pathogenesis

Nucleic Acid or Protein	Associated Virus	SEQ ID NO:	GenBank Accession Nos for cDNA and Protein
T-cell receptor V beta chain	HIV	1-19	

T-cell receptor V-D-J beta 2.1 chain	HIV	20	
β-chimerin (CHN2)	HIV	21-22	NM_004067; NP_004058.1
Malic enzyme 1 (ME1)	HIV and Influenza A	23	BC025246; AAH25246.1
Hypothetical protein XP_174419	HIV and Influenza A	24	
sequence from Chromosome 4q31.3-32	HIV and Influenza A	25-27	
alpha satellite DNA	HIV	28	
LOC253788 and LOC219938; coagulation factor III (F3) and LOC91759	HIV	29	
similar to KOX4 (LOC131880) and LOC166140	HIV	30	
LOC222474 and similar to Rho guanine nucleotide exchange factor 4, isoform a, APC-stimulated guanine nucleotide exchange factor (LOC221178); T-cell receptor beta	HIV	31	
ribosomal protein L7A-like 4 (RPL7AL4) and v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian) (SRC)	HIV	32	
KIAA0564	HIV	33	
alpha satellite DNA; M96 protein	HIV	34	
hypothetical protein similar to G proteins, especially RAP-2A (LOC57826); LOC161005 and osteoblast specific factor 2 (fasciclin I-like; OSF-2)	HIV	35	
Canis familiaris T-cell leukemia translocation-associated (TCTA) gene, aminomethyltransferase (AMT) gene, dystroglycan (DAG1) gene, and bassoon (BSN) gene	Influenza A	36-37	
LIM domain containing preferred translocation partner in lipoma (LPP)	Influenza A	38-48	
sequence between LOC253121 and hyaluronan synthase 2 (HAS2)	Influenza A	49	
Testin 2 and Testin 3 (TES)	Influenza A	50-57	
PTPN1 gene for protein tyrosine phosphatase, non-receptor type 1	Influenza A	58-59	
sequence between LOC149360 and LOC253961	Influenza A	60	
sequence between KIAA1560 and Tectorin beta (TECTB)	Influenza A	61	
Cadherin related 23 (CDH23)	Influenza A	62	BC032581; AAH32581.1
Myeloid/lymphoma or mixed lineage leukemia, translocated to 10 (MMLT10)	Influenza A	63	

exportin 5 (XPO5) and DNA polymerase eta (POLH)	Ebola	64-66	
heterogenous nuclear riboprotein C (C1/C2) (HNRPC)	Ebola	67-75	
alpha-endosulfine pseudogene (ENSAP) and LOC128741	Ebola	76	
LOC222888	Ebola	77	
LOC138421 and zinc finger protein 297B (ZNF297B)	Ebola	78	
sideroflexin 5 (SFXN5)	Ebola	79	AY044437; AAK95826
importin 9 (FLJ10402)	Ebola	80	
T-cell receptor beta	Ebola	81-82	
similar to murine putative transcription factor ZNF131 (LOC135952)	Ebola	83-99	
KIAA1259	Ebola	100-101	AB033085; NP_115572
MURR1 and CCT4	Ebola	102	
FLJ40773 and similar to ribosomal protein L24-like (LOC149360)	Ebola	103	
Testin 2 and 3 (TES)	Ebola	104-107	See above
polybromo 1 (PB1)	Ebola	108	NM_018165.2; NP_060635
DNA damage inducible transcript 3 (DDIT3) and KIAA1887	Ebola	109	
PDZ and LIM domain 1 (elfin) (PDLIM1)	Ebola	110	
LOC284803	Ebola	111-112	
PRO0097 and FLJ31958	Ebola	113	
small inducible cytokine E, member 1 (endothelial monocyte-activating) (SCYE1)	Ebola	114-116	
E3 ubiquitin ligase (SMURF2) and MGC40489	Ebola	117-119	
Ras oncogene family member Rab9	Ebola	118-119	
PRO1617 and retinoblastoma binding protein 1 (RBBP1)	Ebola	120-122	NM_000321; NP_000312.1
region of chromosome 2q12	Ebola	123	
elongation factor for selenoprotein translation	Ebola	124	NM_021937.1 NP_068756.1
Transcription factor SMIF (HSA275986)	Ebola	125-137	
KIAA1026	Ebola	138	
trinucleotide repeat containing 5 (TNRC5)	Ebola	139	
homogentisate 1,2-dioxygenase (HGD)	Ebola	140	
region of chromosome Xq23-24	Ebola	141	
region of chromosome 4p15.3	Ebola	142	
similar to LWamide neuropeptide precursor protein [Hydractinia echinata] (LOC129883)	Ebola	143	

region of chromosome 2q21	Ebola	144	
region of chromosome Xp11.4, including UPS9X	Ebola	145	
LOC221829	Ebola	146	
U3 small nuclear RNA	Ebola	147-154	
integrin, beta 1 (ITGB1)	Ebola	155-158	BC020057; AAH20057.1
acrosomal vesicle protein 1 (ACRV1) and CHK1 checkpoint homolog (CHEK1)	Ebola	159	
prospero-related homeobox 1 (PROX1)	Ebola	160	
FLJ20627 and FLJ12910	Ebola	161-173	
PIN2-interacting protein (PINX1) and SRY (sex-determining region Y)-box 7 (SOX7)	Ebola	174	
LOC131920	Ebola	175	
region of chromosome 13q14	Ebola	176	
neurotrophic tyrosine kinase, receptor, type 3 (NTRK3)	Ebola	177	
TERA protein and FLJ13224	Ebola	178-179	
LOC284260	Ebola	180	
POM (POM121 homolog) and ZP3 fusion (POMZP3)	Ebola	181-182	
DEAD/H box polypeptide 8 (DDX8) and similar to ribosomal protein L29 (cell surface heparin binding protein HIP) (LOC284064)	Ebola	183	
LOC345307 and UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GALNT7)	Ebola	184-186	
Mus musculus 5S rRNA pseudogene (Rn5s-ps1)	Ebola	187	
ribosomal protein L27a pseudogene (RPL27AP) and v-myb myeloblastosis viral oncogene homolog-like 2 (MYBL2)	Ebola	188-192	
Down's syndrome cell adhesion molecule like 1 (DSCAML1)	Ebola	193	
LOC148529	Ebola	194	
Huntingtin-associated protein interacting protein (HAPIP)	Ebola	195	NM_005338.4; NP_005329.3
LOC158525 and similar to RIKEN cDNA 1210001E11 (LOC347366)	Ebola	196-200	
hypothetical protein FLJ12910	Ebola	201-204	
LOC350411	Ebola	205	
allograft inflammatory factor 1 (AIF1) and HLA-B associated transcript 2 (BAT2)	Ebola	206	
C10orf7	Ebola	207	
LOC346658 and LOC340349	Ebola	208-209	
region of chromosome 12q21	Ebola	210	
LOC339248 and FLJ22659	Ebola	211	
SR rich protein DKFZp564B0769 and	Ebola	212	

hypothetical protein MGC14793			
FLJ10439	Ebola	213-214	NM_018093.1; NP_060563.1
cytochrome P450, family 11, subfamily A, polypeptide 1 (CYP11A1) and sema domain, immunoglobulin domain (Ig) and GPI membrane anchor, (semaphoring) 7A	Ebola	215-218	
ribosomal protein S16 (RPS16)	Ebola	219-220	BC004324.2; AAH04324.1
hypothetical protein DKFZp434H0115 and ATP citrate lyase (ACLY)	Ebola	221-222	
calnexin (CANX); protein tyrosine phosphatase, receptor type, K (PTPRK)	Ebola	223-224	
cyclin M2 (CNNM2)	Ebola	225	NM_017649.2; NP_060119.2
AXL receptor tyrosine kinase (AXL)	Ebola	226	BC032229; AAH32229.1
<i>Homo sapiens</i> chromosome 10 open reading frame 3	Ebola	227-228	
<i>Homo sapiens</i> chromosome 10 open reading frame 3 (C10orf3)	Ebola	229-230	
<i>Homo sapiens</i> fer-1-like 3, myoferlin (<i>C. elegans</i>)	Ebola	231-232	NM_013451.; NP_038479.1

Some of the host nucleic acids described in Table 1 and target sequences associated with SEQ ID NOS: 1-227, 229, and 231 encode polypeptides that are receptors or ligands recognized by a particular virus, such as HIV, influenza A, or the Ebola virus. For example, the T-cell receptor V

5 beta and V-D-J beta 2.1 chain polypeptides are part of the T-cell receptor complex that are recognized by certain glycoproteins in the HIV envelope. Other host nucleic acids encode polypeptides that provide an enzymatic function related to a viral life cycle, such as the signaling pathways controlling viral packaging or enzymes involved in viral replications. For example, the β -chimerin rho-GTPase may mediate a cellular signal that initiates or triggers a process leading to 10 passage of an HIV viral particle into the host cell. The data presented herein indicate that Rab9 is involved in pathogen infectivity, for example by interfering with trafficking of proteins and lipids within cells. In particular examples, it is demonstrated that Rab9 is involved in lipid raft formation, and that decreasing functional Rab9 and lipid rafts decreases the ability of pathogens, such as viruses and bacteria, that hijack lipid rafts to bud or be infectious.

15 Still other host nucleic acids participate in the life cycle of a virus. For example, a certain nucleotide sequence of a host nucleic acid, such as a gene within the host genome can be recognized during insertion and integration of a viral genome (reverse transcribed into DNA from the viral RNA genomic template) into the host genome. Viral integration is described in, for example, Coffin *et al.*, *Retroviruses*, Chapter 5.

20 The nucleic acids and proteins disclosed herein can be identified, isolated, and characterized using any number of techniques of molecular biology, including the specific methods and protocols described herein, such as in the examples below. In some examples, the nucleic acids were identified

and isolated using the Lexicon Genetics, Inc. (The Woodlands, TX) "gene trap" technology disclosed in U.S. Pat. Nos: 6,080,576; 6,136,566; 6,207,371; 6,139,833; 6,218,123 and 6,448,000.

Gene trap technology is a powerful method for cloning and identifying functional genes, as it marks a gene with a tag and simultaneously generates a corresponding genetic variation for that 5 particular locus. The method involves introducing into a cell a DNA construct that can monitor and potentially disrupt the transcriptional activity of the region of the cell's genome into which it is inserted. The gene-trap method used to identify the host sequences is disclosed in U.S. Patent No. 6,448,000 (herein incorporated by reference).

Briefly, the gene trap protocol involves infecting a host cell (for example, a cell of a Sup T-1 10 cell line (human), MDCK cells (canine), or Vero cells (monkey)) with a recombinant vector (for example, U3neoSV1, FIG. 1). The recombinant vector includes a selectable marker or other sequence capable of being used to select infected host cells. However, the selectable marker or other sequence does not have a promoter at its 5' end. An exemplary selectable marker is a nucleic acid encoding resistance to an antibiotic (such as neomycin). A summary of the gene trap method is 15 provided in FIGS. 2 and 3. Infection of the host cell is performed in culture under conditions that yield about one copy of the vector per cell. The vector incorporates into the host cell genome adjacent to an active promoter and interrupts or disrupts the transcription of a nucleic acid in the host cell (FIG. 2). The host promoter drives expression of the selectable marker or other sequence on the vector, and infected cells can then be selected. For example, if the vector carries a nucleic acid 20 encoding neomycin resistance, cells can be selected on a medium that contains neomycin or G418, the neomycin analog for mammalian cells, depending on the type of host cell used.

The selected host cells are expanded in culture to form a library of cells that contain 25 randomly disrupted host genes (FIG. 3). An aliquot of the library of cells is exposed to the appropriate virus, such as HIV, influenza A, or Ebola, to determine the effect of the disrupted host sequence on viral infection of the host cells. Host cells that survive the viral infection, or are relatively resistant to such infection (such as those cells that survive for a longer period of time than about at least 50% of the infected cells), can include one or more disrupted genes involved in viral 30 infection. Thus, by using the vector one can decrease viral or pathogenic infection of a host cell or in a subject. Therefore, by identifying these disrupted genes that decreased or otherwise interfered with viral infection of the host cell, candidate sequences are identified that can be used as targets to decrease or inhibit viral infection.

Those host cells that survive viral infection, or are relatively resistant to such infection, are 35 cloned, for example, by limit dilution using a chambered plate or by growth on methylcellulose. The interrupted host nucleic acid is identified using standard molecular biology methods. For example, host DNA can be isolated from the cell and digested using an appropriate restriction enzyme to free the 5' and 3' sequences adjacent the incorporated vector. The isolated DNA fragment can then be amplified, for example using PCR or by introducing the DNA fragment into a bacterial host cell then growing the bacteria. Once isolated, the host nucleic acid can be further characterized and analyzed.

For example, the nucleic acid can be sequenced and compared to other similar nucleic acids.

Methods of using these nucleic acids, and the proteins encoded thereby, are discussed below.

Using these gene trap methods, several host molecules were identified that were previously not known to be involved in viral pathogenesis (SEQ ID NOS: 1-232, Table 1, and target sequences associated with SEQ ID NOS: 1-232). For example, the AMT gene (target sequences associated with SEQ ID NOS: 36 and 37) participates in influenza A infection of host cells. Fragments of host sequences involved in viral infection and pathogenesis can now be identified, even including fragments or sequences that were previously known to be important in the pathogenesis of intracellular pathogens. For example, although the T-cell receptor was previously implicated in HIV infection, the results disclosed herein demonstrate that the T-cell receptor V-D-J beta 2.1 chain (target sequences associated with SEQ ID NO: 20) is involved and in some examples required for HIV infection, and host cells lacking the T-cell receptor V-D-J beta 2.1 chain are unexpectedly highly resistant to HIV infection. Hence the V-D-J beta 2.1 chain is a target for anti-viral therapy at the DNA or polypeptide level, and other pathogenically active subcomponents of other known pathogenic sequences can also be identified with this method.

Examples of these host nucleic acid molecules are target sequences associated with SEQ ID NOS: 1-227, 229, and 231 (including variants, fragments, and fusions thereof) and summarized in Table 1. In addition to these specifically disclosed nucleotide sequences, a host nucleic acid can include nucleotide sequences that are similar to any of the target sequences associated with SEQ ID NOS: 1-227, 229, and 231, such as having at least 70% identity, at least 80% identity, at least 85% identity, at least 90% identity, at least 95% identity, at least 98% identity, or even at least 99% identity to any of the target sequences associated with SEQ ID NOS: 1-227, 229, and 231. The disclosed host nucleic acid sequences, and methods of using them, may comprise, consist, or consist essentially of any of the disclosed nucleic acid sequences shown in SEQ ID NOS: 1-227, 229, and 231, as well as target sequences associated with SEQ ID NOS: 1-227, 229, and 231, or variants or fragments thereof, or sequences that hybridize to the identified sequences under stringent or moderately stringent conditions.

The host nucleic acid molecules also include a fragment of any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, such as a probe or primer as described below.

Host polypeptides corresponding to these nucleic acids also can be used to practice the disclosed methods. In some examples, the polypeptide includes an amino acid sequence that corresponds to a coding sequence of any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, or a target protein sequence associated with SEQ ID NOS: 228, 230, and 232. However, host polypeptides can also include those having similar amino acid sequences, such as polypeptides that are at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, at least 98% identical, or at least 99% identical to the amino acid sequences corresponding to translations of the coding sequence of any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, or a target protein sequence associated with SEQ ID NOS: 228, 230, and 232. For example, the disclosed host polypeptides and methods of using them, may comprise, consist, or consist

essentially of an amino acid sequence corresponding to a translation of the nucleotide sequence in any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, a target protein sequence associated with SEQ ID NOS: 228, 230, and 232, or any of the protein sequences listed in Table 1. Alternatively, the polypeptides include homologous polypeptides from other mammals (for example 5 human, monkeys, and dogs).

The host polypeptide can have an amino acid sequence that varies by one or more conservative substitutions from the amino acid sequences of the proteins encoded by target sequences associated with SEQ ID NOS: 1-227, 229, and 231, or from the target amino acid sequences associated with SEQ ID NOS: 228, 230, and 232. In one example, there is no more than 1, 2, 3, 4, 5, 10 or 10 conservative amino acid substitutions. In another example, there are 1, 2, 3, 4, 5 or 10 conservative amino acid substitutions. The effects of these amino acid substitutions, deletions, or additions on host polypeptides can be assayed, for example, by analyzing the ability of cells transformed with the derivative proteins to resist infection by the corresponding virus.

Also included are fragments of any host polypeptide encoded by any of the target sequences 15 associated with SEQ ID NOS: 1-227, 229, and 231, as well as fragments of the target amino acid sequences associated with SEQ ID NOS: 228, 230, and 232. For example, a protein can include at least 5-500 contiguous amino acids of the protein, such as at least 6-200, at least 6-100, at least 10-100, at least 10-50, or at least 20-50 contiguous amino acids of the protein. A host polypeptide fragment can be at least 5, at least 10, at least 15, at least 25, at least 50, at least 100, at least 200, at 20 least 500, or more amino acids of a polypeptide having an amino acid sequence corresponding to a coding region of the nucleotide sequence in any of the target sequences associated with SEQ ID NOS: 1-227, 229, and 231, or a conservative variant thereof, as well as target amino acid sequences associated with SEQ ID NOS: 228, 230, and 232.

Fragments of a nucleic acid target sequences associated with SEQ ID NOS: 1-227, 229, and 25 231 can include 10-5000 contiguous nucleic acids, such as 12-1000, 12-500, 15-100, or 18-50 contiguous nucleic acids. A host nucleic acid fragment can be at least at least 5, at least 10, at least 15, at least 20, at least 25, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 2000, at least 5000 or more contiguous nucleic acids in any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, or a variant or fusion thereof.

Also included are host nucleic acids that encode the same polypeptide encoded by any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, or a conservative variant of the 30 polypeptide, or a fragment thereof. For example, a host nucleic acid provided by target sequences associated with SEQ ID NOS: 36-37 encodes AMT. A second host nucleic acid also can encode an AMT having the same amino acid sequence as the AMT encoded by target sequences associated with 35 SEQ ID NOS: 36-37, a conservative variant of this AMT, or a fragment thereof, yet this second host nucleic acid can have a different nucleotide sequence than a target sequence associated with SEQ ID NOS: 36-37 due to the degeneracy of the genetic code.

Methods of Using Host Sequences to Decrease Viral Infection

The interaction between a host nucleic acid or polypeptide (such as target sequences associated with SEQ ID NOS: 1-232 and those shown in Table 1) and a virus or viral protein can be decreased or inhibited using the methods provided. Decreasing or inhibiting this interaction can be used to decrease viral infection of a host cell, and/or to decrease symptoms associated with a viral infection in a subject. For example, decreasing or even inhibiting the interaction of a host nucleic acid or polypeptide and a virus can decrease, inhibit, or even prevent infection of a host cell by that virus, or otherwise inhibit the progression or clinical manifestation of the viral infection. In addition, decreasing the interaction of a host nucleic acid or polypeptide and a virus can reduce or alleviate one or more symptoms associated with viral infection, such as a fever.

Several methods can be used to decrease or inhibit the interaction between a viral protein and a host protein or nucleic acid. The viral and host proteins or nucleic acids can be part of an *in vitro* solution, an *in vivo* expression system, or *in situ* with a host tissue or subject. The viral protein can be part of a larger molecule or complex, such as an envelope protein on the envelope of a mature virus or a fragment of a viral envelope. The host protein also can be part of a larger molecule or complex, such as a host polypeptide expressed as part of a fusion protein or contained as one subunit of a larger protein, such as a transport protein, cell receptor, structural protein, or an enzyme. A host nucleic acid can be part of a larger molecule, complex, organism or microorganism such as a host nucleic acid contained within its host genome, a recombinant vector, or a transgenic organism or microorganism (including both extrachromosomal molecules or genomic insertions).

In accordance with the disclosed methods, interaction is decreased or inhibited between a virus or viral protein and more than one (such as 2 or more, such as 3 or more) host nucleic acids or polypeptides. Decreasing or inhibiting the interactions of one or more host nucleic acids or polypeptides with one or more viral proteins can have additive or exponentially increasing effects. For example, it is believed that decreasing the interaction between a host T-cell receptor V-D-J beta 2.1 chain and HIV, or decreasing the activity of a host β -chimerin, within a host cell can enhance the inhibitory effect on HIV infection of that host cell compared to inhibiting the interaction of only one of the host polypeptides. Hence, the methods include interfering with an interaction between the virus or viral protein and more than one of the proteins associated with infection by the same virus.

For example, for infection with HIV, the method could interfere with one, or two or more (such as three or more) of the following: T-cell receptor V beta chain; T-cell receptor V-D-J beta 2.1 chain; β -chimerin (CHN2); malic enzyme 1; Hypothetical protein XP_174419; sequence from Chromosome 4q31.3-32; alpha satellite DNA; LOC253788; LOC219938; coagulation factor III (F3); LOC91759; similar to KOX4 (LOC131880); LOC166140; LOC222474; similar to Rho guanine nucleotide exchange factor 4, isoform a; APC-stimulated guanine nucleotide exchange factor (LOC221178); T-cell receptor beta; ribosomal protein L7A-like 4 (RPL7AL4); v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian) (SRC); KIAA0564; alpha satellite DNA; M96 protein; hypothetical protein similar to G proteins (such as RAP-2A; LOC57826); LOC161005 and osteoblast specific factor 2 (fasciclin I-like).

For Ebola virus, examples of targets include one, or two or more (such as three or more) of the following: exportin 5; DNA polymerase eta (POLH); heterogenous nuclear riboprotein C (C1/C2); alpha-endosulfine pseudogene; LOC128741; LOC222888; LOC138421; zinc finger protein 297B; sideroflexin 5; importin 9; T-cell receptor beta; similar to murine putative transcription factor 5 ZNF131 (LOC135952); KIAA1259; MURR1; CCT4; FLJ40773 and similar to ribosomal protein L24-like (LOC149360); testin 2; testin 3; polybromo 1; DNA damage inducible transcript 3 (DDIT3); KIAA1887; PDZ and LIM domain 1 (elfin) (PDLIM1); LOC284803; PRO0097 and FLJ31958; small inducible cytokine E, member 1 (endothelial monocyte-activating); E3 ubiquitin ligase (SMURF2) and MGC40489; Rab9; PRO1617 and retinoblastoma binding protein 1 (RBBP1); region of 10 chromosome 2q12; elongation factor for selenoprotein translation; transcription factor SMIF (HSA275986); KIAA1026; trinucleotide repeat containing 5; homogentisate 1,2-dioxygenase; region of chromosome Xq23-24; region of chromosome 4p15.3; similar to LWamide neuropeptide precursor protein [Hydractinia echinata] (LOC129883); region of chromosome 2q21; region of chromosome Xp11.4, including UPS9X; LOC221829; U3 small nuclear RNA; integrin, beta 1; acrosomal vesicle 15 protein 1 (ACRV1) and CHK1 checkpoint homolog (CHEK1); prospero-related homeobox 1 (PROX1); FLJ20627 and FLJ12910; PIN2-interacting protein (PINX1) and SRY (sex-determining region Y)-box 7 (SOX7); LOC131920; region of chromosome 13q14; neurotrophic tyrosine kinase, receptor, type 3 (NTRK3); TERA protein; FLJ13224; LOC284260; POM (POM121 homolog) and ZP3 fusion (POMZP3); DEAD/H box polypeptide 8 (DDX8) and similar to ribosomal protein L29 20 (cell surface heparin binding protein HIP) (LOC284064); LOC345307 and UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GALNT7); 5S rRNA pseudogene; ribosomal protein L27a pseudogene (RPL27AP) and v-myb myeloblastosis viral oncogene homolog-like 2 (MYBL2); Down's syndrome cell adhesion molecule like 1; LOC148529; Huntingtin-associated protein interacting protein; LOC158525 and similar to RIKEN cDNA 1210001E11 25 (LOC347366); hypothetical protein FLJ12910; LOC350411; allograft inflammatory factor 1 (AIF1); HLA-B associated transcript 2 (BAT2); C10orf7 ;LOC346658; LOC340349; region of chromosome 12q21; LOC339248; FLJ22659; SR rich protein DKFZp564B0769; hypothetical protein MGC14793; FLJ10439; cytochrome P450, family 11, subfamily A, polypeptide 1; sema domain, immunoglobulin domain (Ig) and GPI membrane anchor, (semaphoring) 7A; ribosomal protein S16; hypothetical 30 protein DKFZp434H0115; ATP citrate lyase; calnexin; protein tyrosine phosphatase, receptor type, K (PTPRK); cyclin M2; AXL receptor tyrosine kinase; *Homo sapiens* chromosome 10 open reading frame 3, mRNA (cDNA clone MGC:3422 IMAGE:3028566); *Homo sapiens* chromosome 10 open reading frame 3 (C10orf3); and *Homo sapiens* fer-1-like 3, myoferlin (*C. elegans*) (FER1L3), transcript variant 1.

35 For influenza, examples of targets include one, or two or more (such as three or more) of the following: T-cell leukemia translocation-associated (TCTA) gene, aminomethyltransferase; dystroglycan; BSN; LIM domain containing preferred translocation partner in lipoma (LPP); sequence between LOC253121 and hyaluronan synthase 2 (HAS2); testin 2; testin 3; PTPN1 gene for protein tyrosine phosphatase, non-receptor type 1; sequence between LOC149360 and LOC253961;

sequence between KIAA1560 and tectorin beta; cadherin related 23; myeloid/lymphoma or mixed lineage leukemia, translocated to 10; malic enzyme 1; hypothetical protein XP_174419; and sequence from chromosome 4q31.3-32.

In examples where a host polypeptide is a cell receptor or part of a cell receptor, decreasing 5 or preventing expression of the polypeptide, or altering the three-dimensional structure of the polypeptide, can reduce or inhibit the interaction between the host cell receptor and a viral protein. Similarly, decreasing, inhibiting or preventing expression of a host ligand polypeptide (or altering the structure of such a ligand) can decrease or inhibit an interaction between the viral protein and the ligand. For example, decreasing or inhibiting expression of one or more enzymes involved in viral 10 pathogenesis, such as those listed in Table 1 and those target sequences associated with SEQ ID NOS: 1-232, can block a component of the viral life cycle, such as blocking a signal pathway leading to transcription or translation of the viral genome, or assembly of viral sub-parts. Decreasing or inhibiting the enzymatic activity of an enzyme (rather than its expression) can have a similar effect.

Altering the nucleotide sequence of a host nucleic acid, for example by targeting disruption 15 of the nucleotide sequence using complementary nucleic acid sequences, can decrease, inhibit or prevent integration of a viral nucleic acid into the host nucleic acid. Methods that can be used to interrupt or alter translation of a host nucleic acid include, but are not limited to, using an antisense RNA, RNAi molecule, or an siRNA that binds to a messenger RNA transcribed by the nucleic acid encoding a host polypeptide as described herein. Decreasing or inhibiting the expression of the host 20 nucleic acid can also alter the course of the disease. In one example, altering the nucleotide sequence of a host gene that is targeted by a virus for viral integration can decrease, inhibit, or even prevent, integration of that virus into the host genome.

A host nucleic acid involved in viral infection, including variants, fusions and fragments thereof, can be used to design agents that bind to a target sequence of that nucleic acid, such as 25 antisense nucleic acids or siRNAs. Such nucleic acid binding agents can be used to decrease or inhibit expression of the nucleic acid, to reduce the incidence of viral infection. For example, an expression vector that transcribes antisense RNA or siRNA that recognizes human β -chimerin mRNA is used to transform cell lines obtained from simians. These transformed cell lines are analyzed for 30 infection by simian immunodeficiency virus (SIV), which is related to HIV. If those cells are resistant to SIV infection, the disrupted gene is identified, sequenced, and compared to the human β -chimerin gene. Sequence similarities between the two genes will offer insight into common molecular mechanisms for infection by HIV and SIV, for example, common structural regions within their respective translated proteins.

A binding agent that recognizes a host nucleic acid involved in viral infection can be used 35 for prophylactic or therapeutic purposes. For example, expression vectors having antisense RNA, RNAi molecules, or siRNA molecules that target a host nucleic acid involved in viral infection, such as β -chimerin, are introduced into the bone marrow of a subject. Uptake of the vector and expression of the antisense RNA, RNAi, or siRNA within cells infected by HIV offers a prophylactic or therapeutic effect by disrupting the β -chimerin genes within those cells, thus decreasing or inhibiting

HIV infection. Similarly, expression vectors including Rab9 antisense RNA, RNAi, or siRNA molecules can be introduced into the bone marrow of a subject. Uptake of the vector and expression of Rab9 antisense RNA, RNAi, or siRNA within cells infected by a pathogen that can hijack a lipid raft, such as HIV or Ebola, offers a prophylactic or therapeutic effect by disrupting the Rab9 genes within those cells, thus decreasing or even inhibiting infection by a pathogen that can hijack a lipid raft. The vector, or other nucleic acid carrying the nucleic acid specific binding agent, is introduced into a subject by any standard molecular biology method and can be included in a composition containing a pharmaceutically acceptable carrier.

Decreasing or inhibiting the interaction between a viral protein and a host protein can 10 decrease or inhibit viral infection. Methods that can be used to decrease an interaction between a viral protein and one or more host proteins (such as at least 2 host proteins, or at least 3 host proteins), include but are not limited to, disrupting expression of a host nucleic acid sequence encoding the host protein, (for example by functionally deleting the coding sequence, such as by a mutation, insertion, or deletion), altering the amino acid sequence or overall shape of the host protein, 15 degrading the host protein, employing an agent that interferes with the viral protein or host protein (such as a specific binding agent, for example an antibody or small molecule), or a combination thereof.

For example, expression of a host protein can occur during transcription or translation of a 20 nucleic acid encoding the host protein, or as a result of post-translational modification of a host protein. Methods that can be used to interrupt or alter transcription of a nucleic acid include, but are not limited to, site-directed mutagenesis, including mutations caused by a transposon or an insertional vector; and providing a DNA-binding protein that binds to the coding region of the host protein, thus blocking or interfering with RNA polymerase or another protein involved in transcription. Various inactive and recombinant DNA-binding proteins, and their effects on transcription, are discussed in 25 Lewin, *Genes VII*. Methods that can be used to interrupt or alter translation of a nucleic acid include, but are not limited to, using an antisense RNA or an siRNA that binds to a messenger RNA transcribed by the nucleic acid encoding the host polypeptide as described herein.

For example, exemplary host T-cell receptor polypeptides are encoded by target sequences 30 associated with SEQ ID NOS: 1-20. Disrupting the expression of a nucleic acid including any target sequence associated with SEQ ID NOS: 1-20 can reduce or prevent production of the corresponding T-cell receptor polypeptide, and without access to the T-cell receptor polypeptide, an HIV virus cannot infect the host cell. Even if expression of the host nucleic acid is not completely blocked or disrupted, virus infection can still be inhibited. For example, interference with a host protein encoded by any target sequence associated with SEQ ID NOS: 1-20 reduces the number of T-cell receptors 35 within that host cell available for recognition by an HIV virus, thus inhibiting HIV infection.

It is shown herein that inhibiting the interaction or activity between host Rab9 and HIV and Ebola using Rab9 siRNA molecules decreases infection of a host cell by the virus compared to the amount of infection in the absence of the siRNA molecules.

Host proteins involved in viral infection, such as those encoded by target cDNA sequences associated with SEQ ID NOS: 1-227, 229, and 231, as well as target sequences associated with SEQ ID NOS: 228, 230, and 232, can be used to generate specific binding agents to those proteins. The specific binding agent can be an anti-protein binding agent, such as a monoclonal or polyclonal antibody. Anti-protein binding agents can provide a prophylactic or therapeutic effect, for example by interfering with viral infection. Assays to determine whether an antibody interferes with viral infection are described herein. Antibodies that recognize a host protein involved in viral infection can prevent a virus or portion thereof (such as a viral protein) from binding to a host protein involved in viral infection. For example, a monoclonal or polyclonal antibody that binds to a V beta T-cell receptor on a cell can block the binding of HIV to that T-cell receptor, thus blocking infection of that cell. Effective amounts of such specific binding agents can be administered alone to a subject, or as part of a pharmaceutical composition, for the treatment of viral infection or as a prophylactic measure prior to the time the subject is exposed to the virus. In another example, specific binding agents that recognize a host protein involved in viral infection, such as β -chimerin or Rab9, can be used can be used to screen for the presence of the host protein, in other cells, tissues or lysates, including a biological sample obtained from a subject.

Host nucleic acids and polypeptides described herein, such as target sequences associated with SEQ ID NOS: 1-232, can be used for prophylactic or therapeutic uses. For example, polypeptides with structures mimicking a protein recognized by a virus can be administered to a subject as a pharmaceutical composition. These polypeptides interact with a virus already infecting that subject, or provide a prophylactic defense mechanism against infection if the subject is at risk of exposure to a virus. For example, polypeptides structurally similar to the T-cell receptor V beta 2.1 chain are recognized by HIV. If such polypeptides are administered to an HIV-positive subject, the viruses already present in the subject interact with those polypeptides in addition to that subject's T-cell receptors, thus inhibiting the rate at which HIV infects T-cells. The administered polypeptides act as "decoys" to block HIV from interacting with T-cell receptors. As another example, an agent that otherwise interferes with the interaction between a virus and a host protein can provide a similar prophylactic effect. For example, a chemical compound or anti-AMT binding agent (such as an antibody) that interferes with the interaction between AMT and an influenza virus (including an enzymatic inhibitor of AMT) provides a prophylactic or therapeutic effect against influenza A infection when provided to a host cell or administered to a host subject.

Additionally, the proteins described herein can be used to screen samples for the presence or absence of a particular antibody. For example, a β -chimerin or Rab9 protein can be used in an ELISA to screen a sample obtained from an individual for the presence of anti- β -chimerin or anti-Rab9 antibodies generated by that individual, such as a blood sample.

Using a method similar to that described for nucleic acid binding agents above, protein binding agents (such as agents that specifically bind β -chimerin, Rab9, or V beta T-cell receptor proteins) can be used to screen cells, individuals or populations for the presence or absence of

polypeptides related to infection (such as HIV, Ebola, or influenza infection), thus providing information about the susceptibility or resistance of that individual or population to viral infection.

The host nucleic acids, proteins, and related specific binding agents described herein can be used as models for the design of anti-viral drugs. For example, the three-dimensional structure of a protein described herein, such as β -chimerin, can be used in computer modeling of chemotherapeutic agents that block the activity of that moiety, for example by binding the protein. As another example, a monoclonal antibody can be used in a competitive binding assay to screen for other compounds that bind the same antigen.

10

Screening for Resistance to Infection

Also provided herein are methods of screening host subjects for resistance to infection by characterizing a nucleotide sequence of a host nucleic acid or the amino acid sequence of a host polypeptide (such as those shown in Table 1, or any target sequence associated with SEQ ID NOS: 1-232).

15

For example, the T-cell receptor V beta 2.1 chain nucleic acid of a subject can be isolated, sequenced, and compared to SEQ ID NO: 20 (or a target sequence associated with SEQ ID NO: 20). The greater the similarity between that subject's V beta 2.1 chain nucleic acid and the sequence shown in SEQ ID NO: 20 (or a target sequence associated with SEQ ID NO: 20), the more susceptible that person is to HIV infection, while a decrease in similarity between that subject's V beta 2.1 chain nucleic acid and SEQ ID NO: 20 (or a target sequence associated with SEQ ID NO: 20), the more resistant that subject can be to HIV infection.

20

In another example, the aminomethyltransferase (AMT) nucleic acid of a subject can be isolated, sequenced, and compared to SEQ ID NOS: 36-37 (or a target sequence associated with SEQ ID NOS: 36-37). The greater the similarity between that subject's AMT nucleic acid and the sequence shown in SEQ ID NOS: 36-37 (or a target sequence associated with SEQ ID NOS: 36-37), the more susceptible that person is to influenza A infection, while a decrease in similarity between that subject's AMT nucleic acid and SEQ ID NOS: 36-37 (or a target sequence associated with SEQ ID NOS: 36-37), the more resistant that subject can be to influenza A infection.

25

In yet another example, the Ras oncogene family member Rab9 nucleic acid of a subject can be isolated, sequenced, and compared to SEQ ID NOS: 118-119 (or a target sequence associated with SEQ ID NOS: 118-119). The greater the similarity between that subject's Rab9 nucleic acid and the sequence shown in SEQ ID NOS: 118-119 (or a target sequence associated with SEQ ID NOS: 118-119), the more susceptible that person is to infection by a pathogen that uses lipid rafts, such as those listed in Table 2, while a decrease in similarity between that subject's Rab9 nucleic acid and SEQ ID NOS: 118-119 (or a target sequence associated with SEQ ID NOS: 118-119), the more resistant that subject may be to infection by a pathogen that uses lipid rafts.

Assessing the genetic characteristics of a population can provide information about the susceptibility or resistance of that population to viral infection. For example, polymorphic analysis of AMT alleles in a particular human population, such as the population of a particular city or

geographic area, can indicate how susceptible that population is to influenza A infection. A higher percentage of AMT alleles substantially similar to SEQ ID NOS: 36-37 (or a target sequence associated with SEQ ID NOS: 36-37) indicates that the population is more susceptible to influenza A infection, while a large number of polymorphic alleles that are substantially different than SEQ ID NOS: 36-37 (or a target sequence associated with SEQ ID NOS: 36-37) indicates that a population is more resistant to influenza A infection. Such information can be used, for example, in making public health decisions about vaccinating susceptible populations.

Transgenic Cells and Non-Human Mammals

10 Transgenic animal models, including recombinant and knock-out animals, can be generated from the host nucleic acids described herein. Exemplary transgenic non-human mammals include, but are not limited to, mice, rats, chickens, cows, and pigs. In certain examples, a transgenic non-human mammal has a knock-out of one or more of the target sequences associated with SEQ ID NOS: 1-35, and has a decreased viral susceptibility, for example infection by HIV. In certain
15 embodiments, a transgenic non-human mammal has a knock-out of any of the target sequences associated with SEQ ID NOS: 36-63, and has a decreased viral susceptibility, for example infection by influenza A. In certain examples, a transgenic non-human mammal has a knock-out of any of the target sequences associated with SEQ ID NOS: 64-232, and has a decreased viral susceptibility, for example infection by Ebola. In certain examples, a transgenic non-human mammal has a knock-out
20 of any target sequence associated with SEQ ID NOS: 118-119, and has a decreased susceptibility to infection by a pathogen that uses a lipid raft, such as those listed in Table 2. Such knock-out animals are useful for reducing the transmission of viruses from animals to humans. In addition, animal viruses that utilize the same targets provided herein can be decreased in the animals.

25 Expression of the sequence used to knock-out or functionally delete the desired gene can be regulated by choosing the appropriate promoter sequence. For example, constitutive promoters can be used to ensure that the functionally deleted gene is never expressed by the animal. In contrast, an inducible promoter can be used to control when the transgenic animal does or does not express the gene of interest. Exemplary inducible promoters include tissue-specific promoters and promoters responsive or unresponsive to a particular stimulus (such as light, oxygen, chemical concentration, such as a tetracycline inducible promoter).

30 For example, a transgenic mouse including an AMT gene (such as a target sequence associated with SEQ ID NOS: 36-37), or a mouse having a disrupted AMT gene, can be examined during exposure to various mammalian viruses related to influenza A. Comparison data can provide insight into the life cycles of influenza and related viruses. Moreover, knock-out animals (such as pigs) that are otherwise susceptible to an infection (for example influenza) can be made to determine the resistance to infection conferred by disruption of the gene.

35 Transgenic pigs having a disrupted human protein tyrosine phosphatase gene can be produced and used as an animal model to determine other types of infections, including viral infections in mammals related to influenza A. A transgenic pig resistant to infection by viruses other

than influenza A is used to demonstrate the relatedness of influenza and those other viruses. Transgenic animals, including methods of making and using transgenic animals, are described in various patents and publication, such as WO 01/43540; WO 02/19811; U.S. Pub. Nos: 2001-0044937 and 2002-0066117; and U.S. Pat. Nos: 5,859,308; 6,281,408; and 6,376,743; and the references cited therein.

Cells including an altered or disrupted host nucleic acid or polypeptide having a role in viral infection (such as a target sequence associated with SEQ ID NOS: 1-232), are resistant to infection by a virus (see Example 2). Such cells may therefore include cells having decreased susceptibility to HIV infection (such as cells having altered or disrupted target sequence associated with SEQ ID NOS: 1-35), Ebola infection (such as cells having altered or disrupted target sequence associated with SEQ ID NOS: 64-232), or influenza A (such as cells having altered or disrupted target sequence associated with SEQ ID NOS: 36-63). For example, cells in which a β -chimerin gene was disrupted using the gene-trap method remain CD4 $^{+}$ after HIV infection and do not produce further detectable HIV virus particles. Thus, disrupting the expression of β -chimerin can confer resistance on the cell to infection by HIV. Additionally, interfering with the activity of β -chimerin, such as contacting a β -chimerin with an enzymatic inhibitor or an anti- β -chimerin binding agent, can confer a similar resistance to HIV infection.

Screening for Agents that Decrease Viral Infection

A host nucleic acid or polypeptide involved in viral infection, such as a target sequence associated with SEQ ID NOS: 1-232, and peptides listed in Table 1, can be used to identify agents that inhibit the binding of a virus or viral protein to a host nucleic acid, a host protein, or another target protein capable of binding to the virus or viral protein. In some examples, a host molecule, such as a host protein or nucleic acid is contacted with a viral molecule, such as a virus or portion thereof, for example as a viral protein. One or more test agents are contacted with the host molecule, the viral molecule, both both molecules, before, during or after contacting the host and viral molecules. Subsequently, it is determined whether binding of the viral molecule to the host molecule is decreased in the presence of the test agent, wherein a decrease in binding is an indication that the test agent decreases the binding of viral protein to the target protein.

In other examples, a cell-based assay is used to identify proteins that decrease viral infection, for example using the yeast two-hybrid system.

For example, the binding of the T-cell receptor V-D-J beta 2.1 chain polypeptide to HIV (or an HIV envelope glycoprotein) can be determined in the presence of a test agent. A decrease in binding activity between the T-cell receptor V-D-J beta 2.1 chain polypeptide and HIV indicates that the test agent decreases the binding of HIV to the T-cell receptor V-D-J beta 2.1 chain, and the agent is a candidate for use as an anti-HIV agent. A decrease in binding activity can be determined by a comparison to a reference standard, such as a binding activity reported in the scientific literature, or to a control. Any suitable compound or composition can be used as a test agent, such as organic or inorganic chemicals, including aromatics, fatty acids, and carbohydrates; peptides, including

monoclonal antibodies, polyclonal antibodies, and other specific binding agents; or nucleic acids. The virus or viral molecule can be obtained from any suitable virus, such as HIV, influenza A, Ebola, and related viruses.

Therapeutic agents identified with the disclosed approaches can be used as lead compounds 5 to identify other agents having even greater antiviral activity. For example, chemical analogs of identified chemical entities, or variant, fragments of fusions of peptide agents, are tested for their ability to decrease viral infection using the disclosed assays. Candidate agents are also tested for safety in animals and then used for clinical trials in animals or humans.

10

Microarrays

The host nucleic acids or proteins disclosed herein having a role in viral infection, such as a target sequence associated with SEQ ID NOS: 1-232, can be used in an array. The array can be a microarray, such as a nucleic acid array that includes probes to different polymorphic alleles of a human AMT gene (for example target sequence associated with SEQ ID NOS: 36-37) or a human 15 Rab9 gene (for example target sequence associated with SEQ ID NOS: 118-119). Kits can be generated, such as diagnostic kits or kits for screening for the presence or absence of a host nucleic acid within a biological sample obtained from a subject or kits for administering an effective amount of a specific binding agent to a subject for a therapeutic or prophylactic purpose.

20

The following examples are provided to illustrate particular features of certain embodiments, but the scope of the claims should not be limited to those features exemplified.

Example 1

Generation of Cells with Increased Resistance to Viral Infection

25 The gene-trap method was used to identify cellular genes needed for viral propagation but whose inactivation is not lethal to the host cell. This was accomplished by using a Moloney murine leukemia virus-derived shuttle vector that encodes for a promoterless neomycin-resistance gene (FIG. 1). This vector integrates into the host genome at transcriptionally active genes, thereby disrupting the host gene but utilizing the host promoter to drive neomycin resistance carried by the vector. The 30 cells are then infected with the desired virus. Cells surviving the viral infection carry an interrupted host gene that is needed during the viral life cycle. Since the construct is a shuttle vector, it can function as a plasmid and can be moved from mammalian to bacterial systems, facilitating subcloning and DNA sequencing. Using this approach, loci involved in, and in some cases required for viral infection, for example by HIV-1 and HIV-2, influenza A and Ebola virus were identified.

35

Tissue culture

Sup-T1 human lymphoblastic leukemia cells were cultured in RPMI-1640 medium supplemented with 10% heat-inactivated fetal calf serum (FCS), penicillin, streptomycin and Fungisome. MDCK normal canine kidney cells were cultured in DMEM supplemented with 10%

fetal bovine serum (FBS), penicillin, streptomycin. Vero African green monkey kidney cells were cultured in DMEM supplemented with 10% FBS, amphotericin B, streptomycin, and Glutamine. All cultures were grown under 5% CO₂. Selection by all media was done in the presence of either 1 mg/ml (Sup-T1 and MDCK cells) or 400 mg/ml G418 (Geneticin; Vero cells).

5

Generation of gene-trapped library of cells

Parental, virus sensitive cells were plated and infected with U3neoSV1 as follows.

- / Retrovirus vectors were obtained from H. Earl Ruley (Department of Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, TN). Stocks of the U3neoSV1 virus were
- 10 prepared as described (Chen *et al.*, Gene trap retroviruses in *Methods in Molecular Genetics* (1994), page 123, herein incorporated by reference).

FIG. 1 illustrates the U3neoSV1 retroviral vector, which contains a promoterless neomycin phosphotransferase gene (*Neo*^R) within the U3 unique sequence of the 5' long terminal repeat (LTR) of MMLV. Additionally, a second mutationally inactivated copy of *neo* is present in the 3' LTR.

- 15 Portions of the MMLV genome were removed to impair replication, and were replaced with the β-lactamase gene which confers ampicillin resistance (*Amp*^R) to *E. coli* as well as an *E. coli* origin of replication (*ori*), flanked by two unique restriction sites for *Bam*HI (position 2570) and *Eco*RI (position 4175). Sites and orientations of primers used for sequence analysis of cloned genomic fragments are indicated by the triangular arrowheads.

- 20 Parental, virus sensitive cells (106 Sup-T1 for HIV, Madin-Darby canine kidney, (MDCK) for influenza A, or Vero cells for Ebola) were plated for 12 hours before infection, after which U3neoSV1 was added at a multiplicity of infection (MOI) of 0.1, as titered by adding 1 ml of diluted stocks to cultured cells in the presence of 4 µg/ml polybrene. The cells were incubated at 37°C for one hour, 10 ml of fresh medium added, and the cells were incubated overnight at 37°C. The next day, the medium was replaced with the appropriate media containing 1 mg/ml G418 and maintained until surviving cells approached confluence, which was usually about two weeks.
- 25

Upon random integration of the U3neoSV1 vector into the host genome, endogenous promoters result in expression of *Neo*^R, while expression of the exons 3' to the site of integration is disrupted. Therefore, only those events occurring at transcriptionally active promoters of non-essential genes are selected.

A pool of the surviving cells, termed a library, including many cells bearing different disrupted genes was then exposed to the pathogen of interest. The resulting Sup-T1 library cells, MDCK library cells, and Vero library cells were infected HIV-1 and HIV-2; the A/PR/8/34 virus reassortant having A/Johannesburg/82/96 glycoproteins (H1N1); and Ebola, respectively, as follows.

- 30
- 35 An aliquot of the cell library was infected with three rounds of HIV-1 and three rounds of HIV-2 (3Bx in BC7 cells), normally a lethal event for Sup T-1 cells (FIG. 4). Approximately 3 × 10⁸ actively growing Sup-T1 library cells were infected with the CXCR4 cytopathic HIV-1 strain LAI at an MOI of 10, approximately 100 fold greater than that normally used for spreading infection in culture. The cells were incubated with the virus for four hours in 2 ml of medium, then grown in bulk

at 10^6 cells/ml for two weeks, at which time G418 was added to a final concentration of 1 mg/ml and the cultures continued for an additional two weeks. The surviving cells were exposed to two further rounds of HIV-1 infection as described above and shown in FIG. 4.

Following HIV-1 infection, surviving cells were incubated 1:100 with BC7 T cells
 5 constitutively expressing the HIV-2 strain 3BX, which was modified to infect regardless of CD4 status, solely using the CXCR4 receptor. Cells were coincubated for two weeks followed by selection with 1 mg/ml G418 (same as FIG. 4, but with HIV-2 instead of HIV-1). The surviving cells were exposed to two further rounds of HIV-2 infection.

The final cell culture was selected using anti-CD4 magnetic microbeads (Miltenyi) and
 10 divided into 2.0 ml cultures containing 1000 cells each. These were then infected with LAI at an MOI of 10. Surviving cells from each culture were subjected to limit dilution, or growth on methylcellulose, and expanded in selection medium. The isolated clones were identified as being CD4 and CXCR4 positive following flow cytometry analysis using standard protocols. Several cell isolates were resistant to further HIV infection with unique expression of CD4 cell surface antigen.

15 For influenza infection, approximately 10^7 actively growing MDCK library cells were washed with phosphate buffered saline (Gibco) and infected with the A/PR/8/34 virus reassortant having A/Johannesburg/82/96 glycoproteins (H1N1) at an MOI of 20-30 in 250 μ l DMEM in a T-25 flask. The cells were incubated with the virus for two hours, and the inoculum was subsequently replaced with DMEM, supplemented with 2% FBS and 1 μ g/ml TPCK trypsin (to cleavage-activate
 20 HA of new progeny virus). The cells were incubated for 18 hours to provide 2-3 rounds of infection. The maintenance medium was removed and replaced with selection medium (DMEM with 10% FBS and 1 mg/ml neomycin) and survivors allowed to expand. The surviving cells were exposed to one additional round of infection as described.

For filovirus infection, vero library cells were infected with either the Gulu 2000 or Zaire
 25 1976 Ebola (EBO) strains, or the Voege 1967 strain of Marburg (MBG) at an MOI of greater than one in T-75 flasks in medium supplemented with 400 mg/ml G418. After a cytopathic effect (CPE) of 4+ was attained (greater than one week), survivors were harvested and reseeded undiluted and at 1:16 and 1:256 dilutions in selection medium. Wells with growth after 10 or more days were reinoculated into T12.5 flasks in selection medium and allowed to expand.

30 Cells surviving Ebola or influenza infection were cloned by either limiting dilution or growth on methylcellulose. The isolates were characterized phenotypically by flow cytometry and the interrupted gene determined by inverse PCR, cloning into BAC, or by the use of the shuttle feature of the vector followed by DNA sequence analysis.

Cloning and Sequencing of Trapped Genes

This example describes the methods used to clone the sequences conferring resistance to the library of cells surviving viral infection. The identified sequences (SEQ ID NOS: 1-227, 229, 231)

encode host proteins that are involved in pathogen infection, and in some cases are required for the infectivity by the pathogen.

Isolation of trapped genes

5 The genomic DNA from actively growing virus-resistant isolates was extracted, prepared, and electroporated into cells as follows. Cellular DNA from actively growing virus-resistant isolates was extracted from one million cells using the QIAamp DNA Blood Mini Kit (Qiagen, Inc.) according to the manufacturer's instructions. Genomic DNA was digested at a final concentration of 150 µg/ml with either *EcoRI* or *BamHI* (New England Biolabs) at 1.5 or 2 units/µl, respectively (see

10 FIG. 1). Digested DNA was ethanol precipitated using oyster glycogen (Sigma) as a carrier, resuspended to a final concentration of 60 ng/µl and ligated using T4 DNA ligase (New England Biolabs). Genomic digestion resulted in the fragmentation of the retrovirus and the genomic DNA. Ligations were subsequently ethanol precipitated in the presence of glycogen, resuspended in 3 µl water and used directly to transform *E. coli*.

15 A 1.5 µl aliquot of each precipitated ligation was added to thawed Genehog cells (Invitrogen) or SURE cells (Stratagene), electroporated using a GenePulser (BioRad) according to the manufacturer's instructions, and plated onto Luria broth (LB) agar (1% tryptone, 0.5% yeast extract, 0.5% NaCl, 2% agar) containing 100 µg/µl carbenicillin (Sigma). Clones were isolated after 24 hours and used to inoculate 3 ml LB containing 100 µg/µl carbenicillin. Plasmid DNA was prepared after
20 overnight growth using the QIAprep Spin Miniprep Kit (QIAGEN, Inc.) according to the manufacturer's instructions and eluted in water.

Sequencing of Shuttle Clones

25 Due to the position of the unique sites in U3neoSV1, *BamHI* digestion facilitates cloning of DNA 3' to the site of integration, while *EcoRI* digestion results in the cloning of genomic DNA 5' to the site of integration. Using oligonucleotides homologous to the U3neoSV1 fragment, the sequence of the disrupted genomic DNA flanking the gene-trap insertion site was determined as follows.

30 Sequencing reactions were performed using the ABI BigDye terminator cycle sequencing kit with reaction products resolved on either an ABI 3100 Genetic Analyzer or an ABI 377 DNA Sequencer (Applied Biosystems, Foster City, CA). Sequences were obtained by using oligonucleotides 5'-ATCTTGTCAATCATGCG (SEQ ID NO: 235) and 5'-GGGTCTGACGCTCATG (SEQ ID NO: 236) for *EcoRI*-generated shuttle clones, or 5'-GATAGGTGCCTCACTG (SEQ ID NO: 237) for *BamHI*-generated shuttle clones.

Sequence analysis

35 Sequences obtained from shuttle clones were analyzed by the Repeatmasker Web Server, available on the Internet at the website for the Department of Molecular Biotechnology, University of Washington, followed by standard nucleotide-nucleotide BLAST (blastn) against the National Center for Biotechnology Information databases, including nr (non-redundant

GenBank+EMBL+DDBJ+PDB sequences), est (expressed sequence tags) and htgs (unfinished High Throughput Genomic Sequences: phases 0, 1 and 2). Additionally, a nucleotide-protein database (blastx) analysis was performed against the nr database.

5 Candidate Host Genes Required for Pathogenesis

Candidate host genes required for the indicated pathogen, which were cloned via the gene-trap method and sequenced, are presented in Table 1 and in SEQ ID NOS: 1-226. The CD4⁺, latently infected, noninfectious HIV-resistant isolates 18B, 18E, 2B, and 2E were used to recover the genes involved in HIV-1 and HIV-2 pathogenesis, influenza A-resistant isolates B1, B3, B5, B6, and B7 10 were used to recover the host genes involved in influenza A pathogenesis, and Ebola-resistant isolates ZV and MV were used to recover the host genes involved in Ebola pathogenesis. Candidate genes can be validated by siRNA and cDNA complementation, as described in Example 3.

In summary, using the U3neoSV1 gene-trap, sixteen HIV-1 and -2 resistant Sup-T1 cell lines, and fifteen influenza A resistant MDCK cell lines were isolated and characterized. Twenty- 15 three EBO-Zaire resistant Vero cell line pools, twenty-four EBO-Gulu resistant pools, and thirty MBG resistant pools were screened. The shuttle-vector design of the U3neoSV1 gene-trap allowed identification of multiple host genes involved in the pathogenesis of HIV-1, HIV-2, influenza A, and Ebola, which are described herein and summarized in Table 1 and sequences provided in SEQ ID NOS: 1-232. Cross-resistance of resistant isolates to multiple pathogens can be quickly examined to 20 reveal common pathways in the viral life cycles.

Example 3

siRNA Molecules Decrease Viral Infection

This example describes methods used to express siRNAs that recognize Rab9 (such as a target sequence associated with SEQ ID NOS: 118-119), AXL (AXL receptor tyrosine kinase; such as a target sequence associated with SEQ ID NO: 226), CHN (beta-chimerin; such as a target sequence associated with SEQ ID NOS: 21-22), KOX (such as a target sequence associated with SEQ ID NO: 30), RBB (retinoblastoma binding protein 1; such as a target sequence associated with SEQ ID NOS: 120-122), KIAA1259; F3 (such as a target sequence associated with SEQ ID NO: 29), and 25 Mselp (mammalian selenium binding protein; such as a target sequence associated with SEQ ID NO: 124).

The following Rab9 siRNA sequences were generated by Dharmacon, RNA Technologies (Lafayette, CO) using chemical synthesis: GGGAAAGAGTTCACTTATGA (SEQ ID NO: 238); TCACAAAGCTTCCAGAACT (SEQ ID NO: 239); GTAACAAGATTGACATAAG (SEQ ID NO: 30 240); and GGAAGTGGATGGACATTIT (SEQ ID NO: 241).

The following AXL (AXL receptor tyrosine kinase) siRNA sequences were generated by Dharmacon, RNA Technologies using chemical synthesis: GGUCAGAGCUGGAGGAUUU (SEQ ID NO: 242); GAAAGAAGGAGACCCGUUA (SEQ ID NO: 243);

CCAAGAAGAACUACAAUGG (SEQ ID NO: 244); and GGAACUGCAUGCUGAAUGA (SEQ ID NO: 245).

5 siRNA sequences were also used that recognized CHN (beta-chimerin); KOX (similar to KOX4 (LOC131880) and LOC166140); RBB (retinoblastoma binding protein 1); KIAA1259; F3 and mammalian selenium binding protein. One skilled in the art will understand that siRNA sequences that recognize other sequences involved in viral infection (such as a target sequence associated with any of SEQ ID NOS: 1-232) can be designed and prepared by commercial entities, such as Dharmacon, RNA Technologies.

10 The four siRNA sequences for each gene (CHN, KOX, RBB, RAB, KIAA1259, F3, ASL and Mselb) were separately pooled. Each of the eight pools of siRNAs, hybridized to its appropriate complement sequence, were used to transfect JC53 (HeLa cells modified to accept HIV), Vero (monkey kidney cells), MDCK (dog kidney cells), or HEK (human kidney cells). All cells were obtained from American Type Culture Collection (ATCC, Mannassas, VA). GFP siRNA sequences were used as a negative control.

15 Cells (20,000 to 250,000) were incubated in serum free media for 24 hours. Cocktails were made by mixing the appropriate duplex siRNAs (50-100 pmoles) with lipofectamine 2000 (4-16 µl) and RNase Inhibitor (1-4 µl) in a solution of Optimem (serum free medium) in a total volume of 200-2000 µl. The lipofectamine was allowed to incubate at room temperature for 5 minutes before the addition of siRNA. Aliquots (50-500 µl) of the cocktail were added to the cells which were 20 incubated at 37°C for 48 hours. The cells were then infected with HIV, Ebola, or influenza and the incubation continued for 3-7 days. Following transfection, several assays were conducted to confirm transfection efficiency, and to determine the resistance of the cells to infection by various agents.

25 Quantitation of p24 levels in HIV infected JC53 cells was determined using the Coulter HIV-1 p24 Antigen Neutralization Kit according to the manufacturers recommendation. As shown in FIG. 5, Rab9 siRNAs and mammalian selenium binding protein siRNAs each decreased HIV infection by about 50% on day 4 post infection (day 7 post addition of siRNA). In addition, HIV infection decreased by about 80-90% in the presence of beta-chimerin siRNAs, KOX (similar to KOX4 (LOC131880) and LOC166140) siRNAs, or retinoblastoma binding protein 1 siRNAs. However, HIV infection did not decrease in the presence of siRNAs that recognize KIAA1259, tissue factor 3, or AXL receptor tyrosine kinase. It is possible that apoptosis is interrupted by the siRNAs, so the cell lives through the infection but still makes virus. It is also possible that the p24 levels are elevated but is not associated with infectious particles.

30 To determine the level of Ebola infection in HEK293 cells transfected with Rab9 or AXL siRNA, the presence of gp1 antigen was determined by using a fluorescent antibody to gp1 envelope protein. Infection by Ebola decreased by at least about 90-95% in the presence of Rab9 siRNA, as compared to the amount of infection in the absence of Rab9 siRNA. Infection by Ebola decreased by at least about 80% in the presence of AXL siRNA, as compared to the amount of infection in the absence of AXL siRNA.

Example 4**Expression of Rab9 siRNA Decreases Lipid Raft Formation**

As described in Example 3, siRNA molecules that recognize Rab9 decrease viral infection.

Rab9 transports late endosomes to trans-golgi. Based on these results, a model is proposed whereby

5 Rab9 plays a role in lipid raft formation (FIG. 6). Lipid rafts are liquid-ordered microdomains enriched in sphingolipids and cholesterol, and are involved in biosynthetic traffic, signal transduction, and endocytosis. Viruses take advantage of ("hijack") rafts for completion of some steps of their replication cycle, such as entry into their cell host, assembly, and budding. Without wishing to be bound to a particular theory, it is proposed that Rab9 trafficks cholesterol, the dynamic glue that

10 holds lipid rafts together. Further evidence for this hypothesis is based on observations of Niemann-Pick type C disease cells. Niemann-Pick type C is a genetic disease that results in accumulation of abnormally high levels of intracellular cholesterol. However, over expression of Rab9 in Niemann-Pick type C disease cells, decreases the level of cholesterol.

Examples of pathogens that hijack lipid rafts include, but are not limited to those shown in

15 Table 2. In the absence of functional Rab9 and lipid rafts (or a decrease in the number of rafts), viruses may not be able to bud or be infectious. Therefore, the use of agents that decrease or inhibit Rab9 expression or activity can be used to decrease infection by other pathogens, as well as toxins such as anthrax, that hijack lipid rafts, such as those shown in Table 2.

20 **Table 2: Pathogens that hijack lipid rafts.**

Bacteria			
Intracellular survival	Toxin binding/oligomerization	Viruses	Protozoa
<i>Campylobacter jujuni</i>	<i>Vibrio cholerae</i>	SV40	<i>Toxoplasma gondii</i>
<i>Legionella pneumophila</i>	<i>Aeromonas hydrophilia</i>	Echovirus 1 and 11	<i>Plasmodium falciparum</i>
<i>Brucella spp</i>	<i>Clostridium spp.</i>	Avian sarcoma and leukosis virus	
<i>FimH and Dr Escherichia coli</i>	<i>Streptococcus pyogenes</i>	Semiliki forest virus	
<i>Salmonella typhimurium</i>	<i>Bacillus anthracis</i>	Ecotropic mouse leukaemia virus	
<i>Shigella flexneri</i>	<i>Bacillus thuringiensis</i>	HTLV-1	
<i>Chlamydia spp.</i>	<i>Helicobacter pylori</i>	HIV-1	
<i>Mycobacterium spp.</i>	<i>Lysteria monocytogenes</i>	Ebola and Marburg viruses	
		Measles virus	
		Herpes Simplex virus	
		Influenza virus	
		Epstein-Barr virus	

This example therefore illustrates that identification of an agent (such as a small molecule or siRNA) that inhibits a particular pathogen can be used to inhibit other pathogens that have a similar mechanism of action.

Example 5
RNAi Molecules

This example describes methods that can be used to decrease or inhibit expression of any of the genes listed in Table 1, or target sequences associated with SEQ ID NOS: 1-232, to decrease viral infection, such as infection by HIV, Ebola, or influenza. Exemplary RNAi compounds are provided for several different genes, such as beta-chimerin receptor tyrosine kinase, retinoblastoma binding protein 1, *Homo sapiens* chromosome 10 open reading frame 3, *Homo sapiens* fer-1-like 3, myoferlin (*C. elegans*), transcript variant 1, *Homo sapiens* chromosome 10 open reading frame 3 (C10orf3), malic enzyme, cadherin related 23, sideroflexin 5, polybromo 1, elongation factor for selenoprotein translation, integrin, beta 1, huntingtin interacting protein 1 and cyclin M2.

One skilled in the art will understand that RNAi molecules can be generated to any of the genes listed in Table 1. Although only 27mers are shown in SEQ ID NOS: 246-845, this disclosure is not limited to RNAi compounds of a particular length. An RNAi molecule can be any length, such as at least about about 25 nucleotides, or even as many as 400 nucleotides. One skilled in the art will also understand that RNAi sequences that recognize other sequences involved in viral infection (such as a target sequence associated with any of SEQ ID NOS: 1-232) can be designed and prepared by commercial entities, such as Sequitur, Inc. (Natick, MA).

Using the methods described in Example 3, the disclosed RNAi compounds are used to decrease viral infection. For example, a 27mer RNAi compound shown in any of SEQ ID NOS: 246-845 is incubated with its reverse complement, allowing hybridization of the two molecules. In particular examples, two or more, such as three or more, 27mer RNAi compounds are transfected into a cell. This duplex molecule is contacted with a cell, such as a cell of a subject in whom decreased viral infection is desired, under conditions that allow the duplex to enter the cell.

25

Example 6

Disruption of Gene Expression

This example describes methods that can be used to disrupt expression of a host gene, such as those shown in Table 1 and target sequences associated with SEQ ID NOS: 1-232, and thereby decrease activity of the proteins encoded by these sequences. Such methods are useful when it is desired to decrease or inhibit viral infection. In a particular example, disrupted expression of at least one target sequence associated with SEQ ID NOS: 1-232 in a host cell is used to treat a subject having a viral infection, or susceptible to a viral infection. Methods useful for disrupting gene function or expression are the use of antisense oligonucleotides, siRNA molecules (see Example 3), RNAi molecules (see Example 5), ribozymes, and triple helix molecules. Techniques for the production and use of such molecules are well known to those of skill in the art.

Antisense Methods

To design antisense oligonucleotides, a host mRNA sequence is examined. Regions of the sequence containing multiple repeats, such as TTTTTTTT, are not as desirable because they will lack specificity. Several different regions can be chosen. Of those, oligos are selected by the following characteristics: those having the best conformation in solution; those optimized for hybridization 5 characteristics; and those having less potential to form secondary structures. Antisense molecules having a propensity to generate secondary structures are less desirable.

Plasmids including antisense sequences that recognize one or more of the target sequences associated with SEQ ID NOS: 1-232 (such as a sequence that encodes a protein listed in Table 1) can be generated using standard methods. For example, cDNA fragments or variants coding for a host 10 protein involved in viral infection are PCR amplified. The nucleotides are amplified using Pfu DNA polymerase (Stratagene) and cloned in antisense orientation a vector, such as pcDNA vectors (InVitrogen, Carlsbad, CA). The nucleotide sequence and orientation of the insert can be confirmed by sequencing using a Sequenase kit (Amersham Pharmacia Biotech).

Generally, the term "antisense" refers to a nucleic acid capable of hybridizing to a portion of 15 a host RNA sequence (such as mRNA) by virtue of some sequence complementarity. The antisense nucleic acids disclosed herein can be oligonucleotides that are double-stranded or single-stranded, RNA or DNA or a modification or derivative thereof, which can be directly administered to a cell, or which can be produced intracellularly by transcription of exogenous, introduced sequences.

Antisense nucleic acids are polynucleotides, and can be oligonucleotides (ranging from 20 about 6 to about 100 oligonucleotides). In one example, an antisense polynucleotide recognizes one or more of the target nucleic acid sequences associated with SEQ ID NOS: 1-227, 229, or 231. In specific examples, the oligonucleotide is at least 10, 15, or 100 nucleotides, or a polynucleotide of at least 200 nucleotides. However, antisense nucleic acids can be much longer. The nucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, and can include other appending 25 groups such as peptides, or agents facilitating transport across the cell membrane (Letsinger *et al.*, *Proc. Natl. Acad. Sci. USA* 1989, 86:6553-6; Lemaitre *et al.*, *Proc. Natl. Acad. Sci. USA* 1987, 84:648-52; WO 88/09810) or blood-brain barrier (WO 89/10134), hybridization triggered cleavage agents (Krol *et al.*, *BioTechniques* 1988, 6:958-76) or intercalating agents (Zon, *Pharm. Res.* 5:539-49, 1988).

An antisense polynucleotide (including oligonucleotides) that recognizes one or more of the target sequences associated with SEQ ID NOS: 1-227, 229, or 231, can be modified at any position 30 on its structure with substituents generally known in the art. For example, a modified base moiety can be 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N-6-sopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, methoxyarninomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-

oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-S-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, and 2,6-diaminopurine.

An antisense polynucleotide that recognizes one or more of the target sequences associated with SEQ ID NOS: 1-227, 229, or 231, can include at least one modified sugar moiety such as arabinose, 2-fluoroarabinose, xylose, and hexose, or a modified component of the phosphate backbone, such as phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, or a formacetal or analog thereof.

In a particular example, an antisense polynucleotide that recognizes one or more of the target sequences associated with SEQ ID NOS: 1-227, 229, or 231 is an α -anomeric oligonucleotide. An α -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gautier *et al.*, *Nucl. Acids Res.* 15:6625-41, 1987). The oligonucleotide can be conjugated to another molecule, such as a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent. Oligonucleotides can include a targeting moiety that enhances uptake of the molecule by host cells. The targeting moiety can be a specific binding molecule, such as an antibody or fragment thereof that recognizes a molecule present on the surface of the host cell.

Polynucleotides disclosed herein can be synthesized by standard methods, for example by use of an automated DNA synthesizer. As examples, phosphorothioate oligos can be synthesized by the method of Stein *et al.* (*Nucl. Acids Res.* 1998, 16:3209), methylphosphonate oligos can be prepared by use of controlled pore glass polymer supports (Sarin *et al.*, *Proc. Natl. Acad. Sci. USA* 85:7448-51, 1988). In a specific example, antisense oligonucleotide that recognizes one or more of the target sequences associated with SEQ ID NOS: 1-227, 229, or 231 includes catalytic RNA, or a ribozyme (see WO 90/11364, Sarver *et al.*, *Science* 247:1222-5, 1990). In another example, the oligonucleotide is a 2'-0-methylribonucleotide (Inoue *et al.*, *Nucl. Acids Res.* 15:6131-48, 1987), or a chimeric RNA-DNA analogue (Inoue *et al.*, *FEBS Lett.* 215:327-30, 1987).

The antisense polynucleic acids disclosed herein include a sequence complementary to at least a portion of an RNA transcript of a gene, such as a target sequence associated with SEQ ID NOS: 1-227, 229, or 231. However, absolute complementarity, although advantageous, is not required. A sequence can be complementary to at least a portion of an RNA, meaning a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation can be assayed. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

The relative ability of polynucleotides (such as oligonucleotides) to bind to complementary strands is compared by determining the T_m of a hybridization complex of the poly/oligonucleotide and its complementary strand. The higher the T_m the greater the strength of the binding of the hybridized strands. As close to optimal fidelity of base pairing as possible achieves optimal 5 hybridization of a poly/oligonucleotide to its target RNA.

The amount of antisense nucleic acid that is effective in the treatment of a particular disease or condition (the therapeutically effective amount) depends on the nature of the disease or condition, and can be determined by standard clinical techniques. For example, it can be useful to use 10 compositions to achieve sustained release of an antisense nucleic acid, for example an antisense molecule that recognizes one or more target sequences associated with SEQ ID NOS: 1-227, 229, or 231. In another example, it may be desirable to utilize liposomes targeted via antibodies to specific cells.

As an alternative to antisense inhibitors, catalytic nucleic acid compounds, such as 15 ribozymes or anti-sense conjugates, can be used to inhibit gene expression. Ribozymes can be synthesized and administered to the subject, or can be encoded on an expression vector, from which the ribozyme is synthesized in the targeted cell (as in WO 9523225, and Beigelman *et al. Nucl. Acids Res.* 1995, 23:4434-42). Examples of oligonucleotides with catalytic activity are described in WO 9506764. Conjugates of antisense with a metal complex, such as terpyridylCu (II), capable of 20 mediating mRNA hydrolysis, are described in Bashkin *et al. (Appl. Biochem Biotechnol. 54:43-56, 1995).*

Ribozymes

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of 25 RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by a endonucleolytic cleavage. Methods of using ribozymes to decrease or inhibit RNA expression are known in the art. An overview of ribozymes and methods of their use is provided in Kashani-Sabet (*J. Investig. Dermatol. Symp. Proc.*, 7:76-78, 2002).

Ribozyme molecules include one or more sequences complementary to the target host 30 mRNA and include the well-known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246, herein incorporated by reference).

A ribozyme gene directed against any of the target sequences associated with SEQ ID NOS: 1-227, 229, or 231 can be delivered to a subject endogenously (where the ribozyme coding gene is transcribed intracellularly) or exogenously (where the ribozymes are introduced into a cell, for 35 example by transfection). Methods describing endogenous and exogenous delivery are provided in Marschall *et al. (Cell Mol. Neurobiol. 14:523-38, 1994).*

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites that include the following sequence: GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and ribonucleotides

corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

For example, a plasmid that contains a ribozyme gene directed against a β -chimerin rho-GTPase, placed behind a promoter, can be transfected into the cells of a subject, for example a subject susceptible to HIV infection. Expression of this plasmid in a cell will decrease or inhibit β -chimerin rho-GTPase RNA expression in the cell. In another example, a plasmid that contains a ribozyme gene directed against Rab9 placed behind a promoter, can be transfected into the cells of a subject, for example a subject susceptible to infection by a pathogen that utilizes lipid rafts, such as Ebola. Expression of this plasmid in a cell will decrease or inhibit Rab9 RNA expression in the cell. Other examples of using ribozymes to decrease or inhibit RNA expression can be found in WO 01/83754 (herein incorporated by reference).

15

Triple helix molecules

Nucleic acid molecules used in triplex formation should be single stranded and composed of deoxynucleotides. The base composition of these oligonucleotides is ideally designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, which will result in TAT and CGC+ triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, for example, contain a stretch of guanine residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation may be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with one strand of a duplex first and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

Example 7

35

Methods of Treatment

When the activity of a host cell protein or nucleic acid involved in viral infection is decreased by prematurely downregulating their levels of expressing using antisense molecules, a reduction in viral infection can be achieved. Antisense oligonucleotides, RNAi molecules, ribozymes, and siRNA molecules that recognize a host nucleic acid involved in viral infection

(Example 6) can therefore be used to disrupt cellular expression of a host protein involved in viral infection. The disclosed antisense, ribozyme, RNAi molecules and siRNA molecules can be administered to a subject alone, or in combination with other therapeutic agents such as anti-viral compounds.

5 A subject susceptible to or suffering from a viral infection, wherein decreased amounts of infection by the virus is desired, can be treated with a therapeutically effective amount of antisense, ribozyme, RNAi molecule or siRNA molecule (or combinations thereof) that recognizes a host sequence involved in viral infection, such as those shown in Table 1 or target sequences associated with SEQ ID NOS: 1-232. After the antisense, ribozyme, RNAi molecule or siRNA molecule has
10 produced an effect (a decreased level of viral infection is observed, or symptoms associated with viral infection decrease), for example after 24-48 hours, the subject can be monitored for diseases associated with viral infection.

Similarly, other agents, such as an antibody that recognizes a host protein involved in viral infection and prevents the protein from interacting with a viral protein, can also be used to decrease
15 or inhibit viral infection. Other exemplary agents are those identified using the methods described in the Examples below. These agents, such as antibodies, peptides, nucleic acids, organic or inorganic compounds, can be administered to a subject in a therapeutically effective amount. After the agent has produced an effect (a decreased level of viral infection is observed, or symptoms associated with viral infection decrease), for example after 24-48 hours, the subject can be monitored for
20 diseases associated with viral infection.

The treatments disclosed herein can also be used prophylactically, for example to inhibit or prevent a viral infection. Such administration is indicated where the treatment is shown to have utility for treatment or prevention of the disorder. The prophylactic use is indicated in conditions known or suspected of progressing to disorders associated with a viral infection.

25

Example 8 **Recombinant Expression**

With the disclosed host sequences involved in viral infection, native and variant sequences can be generated. Expression and purification by standard laboratory techniques of any variant, such
30 as a polymorphism, mutant, fragment or fusion of a sequence involved in viral infection, such as a target sequence associated with SEQ ID NOS: 1-232, is enabled. One skilled in the art will understand that the sequences involved in viral infection, as well as variants thereof, can be produced recombinantly in any cell or organism of interest, and purified prior to use.

Methods for producing recombinant proteins are well known in the art. Therefore, the scope
35 of this disclosure includes recombinant expression of any host protein or variant or fragment thereof involved in viral infection. For example, see U.S. Patent No: 5,342,764 to Johnson *et al.*; U.S. Patent No: 5,846,819 to Pausch *et al.*; U.S. Patent No: 5,876,969 to Fleer *et al.* and Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989, Ch. 17, herein incorporated by reference).

Briefly, partial, full-length, or variant cDNA sequences that encode for a protein involved in viral infection, such as a target sequence associated with SEQ ID NOS: 1-232, can be ligated into an expression vector, such as a bacterial expression vector. Proteins or peptides can be produced by placing a promoter upstream of the cDNA sequence. Examples of promoters include, but are not limited to *lac*, *trp*, *tac*, *trc*, major operator and promoter regions of phage lambda, the control region of fd coat protein, the early and late promoters of SV40, promoters derived from polyoma, adenovirus, retrovirus, baculovirus and simian virus, the promoter for 3-phosphoglycerate kinase, the promoters of yeast acid phosphatase, the promoter of the yeast alpha-mating factors and combinations thereof.

10 Vectors suitable for the production of intact proteins include pKC30 (Shimatake and Rosenberg, 1981, *Nature* 292:128), pKK177-3 (Amann and Brosius, 1985, *Gene* 40:183) and pET-3 (Studiar and Moffatt, 1986, *J. Mol. Biol.* 189:113). A DNA sequence can be transferred to other cloning vehicles, such as other plasmids, bacteriophages, cosmids, animal viruses and yeast artificial chromosomes (YACs) (Burke *et al.*, 1987, *Science* 236:806-12). These vectors can be introduced into a variety of hosts including somatic cells, and simple or complex organisms, such as bacteria, fungi (Timberlake and Marshall, 1989, *Science* 244:1313-7), invertebrates, plants (Gasser and Fraley, 1989, *Science* 244:1293), and mammals (Pursel *et al.*, 1989, *Science* 244:1281-8), that are rendered transgenic by the introduction of the heterologous cDNA.

20 For expression in mammalian cells, a cDNA sequence, such as a coding sequence of any target sequence associated with SEQ ID NOS: 1-227, 229, or 231, can be ligated to heterologous promoters, such as the simian virus SV40, promoter in the pSV2 vector (Mulligan and Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072-6), and introduced into cells, such as monkey COS-1 cells (Gluzman, 1981, *Cell* 23:175-82), to achieve transient or long-term expression. The stable integration of the chimeric gene construct may be maintained in mammalian cells by biochemical selection, such as neomycin (Southern and Berg, 1982, *J. Mol. Appl. Genet.* 1:327-41) and mycophenoic acid (Mulligan and Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072-6).

25 The transfer of DNA into eukaryotic, such as human or other mammalian cells is a conventional technique. The vectors are introduced into the recipient cells as pure DNA (transfection) by, for example, precipitation with calcium phosphate (Graham and vander Eb, 1973, *Virology* 52:466) strontium phosphate (Brash *et al.*, 1987, *Mol. Cell Biol.* 7:2013), electroporation (Neumann *et al.*, 1982, *EMBO J.* 1:841), lipofection (Felgner *et al.*, 1987, *Proc. Natl. Acad. Sci. USA* 84:7413), DEAE dextran (McCuthan *et al.*, 1968, *J. Natl. Cancer Inst.* 41:351), microinjection (Mueller *et al.*, 1978, *Cell* 15:579), protoplast fusion (Schafner, 1980, *Proc. Natl. Acad. Sci. USA* 77:2163-7), or pellet guns (Klein *et al.*, 1987, *Nature* 327:70). Alternatively, the cDNA can be 30 introduced by infection with virus vectors, for example retroviruses (Bernstein *et al.*, 1985, *Gen. Engrg.* 7:235) such as adenoviruses (Ahmad *et al.*, *J. Virol.* 57:267, 1986) or Herpes (Spaete *et al.*, *Cell* 30:295, 1982).

Pharmaceutical Compositions and Modes of Administration

Various delivery systems for administering the therapies disclosed herein are known, and include encapsulation in liposomes, microparticles, microcapsules, expression by recombinant cells, receptor-mediated endocytosis (Wu and Wu, *J. Biol. Chem.* 1987, 262:4429-32), and construction of therapeutic nucleic acids as part of a retroviral or other vector. Methods of introduction include, but are not limited to, topical, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, and oral routes. The compounds can be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (for example, oral mucosa, rectal, vaginal and intestinal mucosa, etc.) and can be administered together with other biologically active agents. Administration can be systemic or local.

Pharmaceutical compositions can be delivered locally to the area in need of treatment, for example by topical application.

Pharmaceutical compositions are disclosed that include a therapeutically effective amount of an RNA, DNA, antisense molecule, ribozyme, RNAi molecule, siRNA molecule, specific-binding agent, or other therapeutic agent, alone or with a pharmaceutically acceptable carrier. Furthermore, the pharmaceutical compositions or methods of treatment can be administered in combination with (such as before, during, or following) other therapeutic treatments, such as other antiviral agents.

Delivery systems

The pharmaceutically acceptable carriers useful herein are conventional. *Remington's Pharmaceutical Sciences*, by Martin, Mack Publishing Co., Easton, PA, 15th Edition (1975), describes compositions and formulations suitable for pharmaceutical delivery of the therapeutic agents herein disclosed. In general, the nature of the carrier will depend on the mode of administration being employed. For instance, parenteral formulations usually include injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, sesame oil, glycerol, ethanol, combinations thereof, or the like, as a vehicle. The carrier and composition can be sterile, and the formulation suits the mode of administration. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example sodium acetate or sorbitan monolaurate.

The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. For solid compositions (for example powder, pill, tablet, or capsule forms), conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, sodium saccharine, cellulose, magnesium carbonate, or magnesium stearate. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides.

Embodiments of the disclosure including medicaments can be prepared with conventional pharmaceutically acceptable carriers, adjuvants and counterions as would be known to those of skill in the art.

The amount of therapeutic agent effective in decreasing or inhibiting viral infection can 5 depend on the nature of the virus and its associated disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* assays can be employed to identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each subject's circumstances. Effective doses can be extrapolated 10 from dose-response curves derived from *in vitro* or animal model test systems.

The disclosure also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by 15 the agency of manufacture, use or sale for human administration. Instructions for use of the composition can also be included.

Administration of Nucleic Acids

In an example in which a nucleic acid is employed to reduce viral infection, such as an 20 antisense, RNAi molecule, or siRNA molecule, the nucleic acid can be delivered intracellularly (for example by expression from a nucleic acid vector or by receptor-mediated mechanisms), or by an appropriate nucleic acid expression vector which is administered so that it becomes intracellular, for example by use of a retroviral vector (see U.S. Patent No. 4,980,286), or by direct injection, or by use of microparticle bombardment (such as a gene gun; Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, or by administering it in linkage to a homeobox-like peptide 25 which is known to enter the nucleus (for example Joliot *et al.*, *Proc. Natl. Acad. Sci. USA* 1991, 88:1864-8). The present disclosure includes all forms of nucleic acid delivery, including synthetic oligos, naked DNA, plasmid and viral, integrated into the genome or not.

This example describes *in vitro* methods that can be used to screen test agents for their ability to interfere with or even inhibit viral infection of a host cell. As disclosed in the Examples above, the disclosed host proteins (such as those listed in Table 1 and the target protein sequences associated with SEQ ID NOS: 1-232, as well as variants, fragments, and fusions thereof) are involved 5 in viral infection (such as infection by HIV, Ebola, and influenza A), and the host protein/viral protein interaction is a component in the ability of a virus to infect a cell. Therefore, screening assays can be used to identify and analyze agents that decrease or interfere with this interaction. For example, the following assays can be used to identify agents that interfere with the interaction of the disclosed host proteins (such as those listed in Table 1 and the target protein sequences associated 10 with SEQ ID NOS: 1-232) with a viral protein sequence. However, the present disclosure is not limited to the particular methods disclosed herein.

Agents identified via the disclosed assays can be useful, for example, in decreasing or even inhibiting viral infection by more than an amount of infection in the absence of the agent, such as a decrease of at least about 10%, at least about 20%, at least about 50%, or even at least about 90%. 15 This decrease in viral infection can serve to ameliorate symptoms associated with viral infection, such as fever. Assays for testing the effectiveness of the identified agents, are discussed below.

Exemplary test agents include, but are not limited to, any peptide or non-peptide composition in a purified or non-purified form, such as peptides made of D-and/or L-configuration amino acids (in, for example, the form of random peptide libraries; see Lam *et al.*, *Nature* 354:82-4, 20 1991), phosphopeptides (such as in the form of random or partially degenerate, directed phosphopeptide libraries; see, for example, Songyang *et al.*, *Cell* 72:767-78, 1993), antibodies, and small or large organic or inorganic molecules. A test agent can also include a complex mixture or “cocktail” of molecules.

The basic principle of the assay systems used to identify agents that interfere with the 25 interaction between a host protein, such as those listed in Table 1 and the target protein sequences associated with SEQ ID NOS: 1-232, and its viral protein binding partner or partners, involves preparing a reaction mixture containing the host protein and a viral protein under conditions and for a time sufficient to allow the two proteins to interact and bind, thus forming a complex. In order to test an agent for inhibitory activity, the reaction is conducted in the presence and absence of the test 30 agent. The test agent can be initially included in the reaction mixture, or added at a time subsequent to the addition of a host protein and a viral protein. Controls are incubated without the test agent or with a placebo. Exemplary controls include agents known not to bind to viral or host proteins. The formation of any complexes between the host protein and the viral protein is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test 35 agent, indicates that the agent interferes with the interaction of the host protein and the viral protein, and is therefore possibly an agent that can be used to decrease viral infection.

The assay for agents that interfere with the interaction of host and viral proteins can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring the host protein or the viral protein onto a solid phase and detecting complexes anchored on the solid

phase at the end of the reaction. In some examples, the method further involves quantitating the amount of complex formation or inhibition. Exemplary methods that can be used to detect the presence of complexes, when one of the proteins is labeled, include ELISA, spectrophotometry, flow cytometry, and microscopy. In homogeneous assays, the entire reaction is performed in a liquid 5 phase. In either method, the order of addition of reactants can be varied to obtain different information about the agents being tested. For example, test agents that interfere with the interaction between the proteins, such as by competition, can be identified by conducting the reaction in the presence of the test agent, for example by adding the test agent to the reaction mixture prior to or simultaneously with the host protein and viral protein. On the other hand, test agents that disrupt 10 preformed complexes, such as agents with higher binding constants that displace one of the proteins from the complex, can be tested by adding the test agent to the reaction mixture after complexes have been formed. The various formats are described briefly below.

Once identified, test agents found to inhibit or decrease the interaction between a host protein and a viral protein can be formulated in therapeutic products (or even prophylactic products) 15 in pharmaceutically acceptable formulations, and used for specific treatment or prevention of a viral disease, such as HIV, Ebola, or influenza A.

Heterogeneous assay system

In a heterogeneous assay system, one binding partner, either the host protein (such as those 20 listed in Table 1 and target protein sequences associated with SEQ ID NOS: 1-232) or the viral protein (such as an HIV, Ebola, or influenza A virus preparation) is anchored onto a solid surface (such as a microtiter plate), and its binding partner, which is not anchored, is labeled, either directly or indirectly. Exemplary labels include, but are not limited to, enzymes, fluorophores, ligands, and radioactive isotopes. The anchored protein can be immobilized by non-covalent or covalent 25 attachments. Non-covalent attachment can be accomplished simply by coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody (such as a monoclonal antibody) specific for the protein can be used to anchor the protein to the solid surface. The surfaces can be prepared in advance and stored.

To conduct the assay, the binding partner of the immobilized species is added to the coated 30 surface with or without the test agent. After the reaction is complete, unreacted components are removed (such as by washing) and any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the binding partner was pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the binding partner is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; for example by using a 35 labeled antibody specific for the binding partner (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds which inhibit complex formation or which disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test agent, the reaction products separated from unreacted components, and complexes detected; for example by using an immobilized antibody specific for one binding partner to anchor any complexes formed in solution, and a labeled antibody specific for the other binding partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test agents which inhibit complex or which disrupt preformed complexes can be identified.

Homogenous assays

In an alternate example, a homogeneous assay can be used. In this method, a preformed complex of the host protein and the viral protein is prepared in which one of the proteins is labeled, but the signal generated by the label is quenched due to complex formation (for example, see U.S. Pat. No. 4,109,496 by Rubenstein which utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the binding partners from the preformed complex will result in the generation of a signal above background. In this way, test agents that disrupt host protein-viral protein interactions are identified.

Immobilization of Proteins

In a particular example, a host protein involved in viral infection (such as those listed in Table 1 and the target protein sequences associated with SEQ ID NOS: 1-232) can be prepared for immobilization using recombinant DNA techniques. For example, a coding region of a protein listed in Table 1, or any target sequence associated with SEQ ID NOS: 1-232, can be fused to a glutathione-S-transferase (GST) gene using the fusion vector pGEX-5X-1, in such a manner that its binding activity is maintained in the resulting fusion protein. The viral protein (such as an Ebola, HIV, or influenza A protein or viral preparation) can be purified and used to raise a monoclonal antibody, using methods routinely practiced in the art and described above. This antibody can be labeled with the radioactive isotope ¹²⁵I using methods routinely practiced in the art.

In a heterogeneous assay, for example, the GST-host fusion protein can be anchored to glutathione-agarose beads. The viral protein preparation can then be added in the presence or absence of the test agent in a manner that allows interaction and binding to occur. At the end of the reaction period, unbound material can be washed away, and the labeled monoclonal antibody can be added to the system and allowed to bind to the complexed binding partners. The interaction between the host protein and the viral protein can be detected by measuring the amount of radioactivity that remains associated with the glutathione-agarose beads. A successful inhibition of the interaction by the test compound will result in a decrease in measured radioactivity.

Alternatively, the GST-host fusion protein and the viral protein can be mixed together in liquid in the absence of the solid glutathione agarose beads. The test agent can be added either during or after the binding partners are allowed to interact. This mixture can then be added to the glutathione-agarose beads and unbound material is washed away. Again, the extent of inhibition of

the binding partner interaction can be detected by adding the labeled antibody and measuring the radioactivity associated with the beads.

In another example, these same techniques can be employed using peptide fragments that correspond to the binding domains of the host protein and the viral protein, respectively, in place of one or both of the full length proteins. Any number of methods routinely practiced in the art can be used to identify and isolate the protein's binding site. These methods include, but are not limited to, mutagenesis of one of the genes encoding the proteins and screening for disruption of binding in a co-immunoprecipitation assay. Compensating mutations in a host gene can be selected. Sequence analysis of the genes encoding the respective proteins will reveal the mutations that correspond to the 5 region of the protein involved in interactive binding. Alternatively, one protein can be anchored to a solid surface using methods described in above, and allowed to interact with and bind to its labeled binding partner, which has been treated with a proteolytic enzyme, such as trypsin. After washing, a short, labeled peptide comprising the binding domain may remain associated with the solid material, which can be isolated and identified by amino acid sequencing. Also, once the gene coding for the 10 for the cellular or extracellular protein is obtained, short gene segments can be engineered to express peptide fragments of the protein, which can then be tested for binding activity and purified or synthesized.

For example, a host protein can be anchored to a solid material as described above by making a GST-host protein fusion protein and allowing it to bind to glutathione agarose beads. The 15 viral protein can be labeled with a radioactive isotope, such as ^{35}S , and cleaved with a proteolytic enzyme such as trypsin. Cleavage products can then be added to the anchored GST-host protein fusion protein and allowed to bind. After washing away unbound peptides, labeled bound material, representing the cellular or extracellular protein binding domain, can be eluted, purified, and analyzed for amino acid sequence. Peptides so identified can be produced synthetically or fused to appropriate 20 facilitative proteins using recombinant DNA technology.

Example 11

Cell-Based Screening Assay for Agents that Decrease Viral Infection

This example describes methods using intact cells that can be used to screen test agents for 25 their ability to interfere with or even inhibit viral infection of a host cell. For example, a yeast two-hybrid assay or the inverse two-hybrid assay method of Schreiber and coworkers (*Proc. Natl. Acad. Sci., USA* 94:13396, 1997) is used to screen for an agent that disrupts the association between a host protein (such as those listed in Table 1, proteins encoded by any target sequence associated with SEQ ID NOS: 1-227, 229, and 231, and any target sequence associated with SEQ ID NOS: 229, 230, and 30 232) and a viral protein (such as HIV, Ebola, or influenza A virus). Similar to Example 10, therapeutic agents identified by these approaches are tested for their ability to decrease or inhibit 35 infection of a host cell, such as a human cell, by HIV, Ebola, or influenza A.

In one example, the yeast two-hybrid system is used to identify anti-viral agents. One version of this system has been described (Chien *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:9578-82,

1991) and is commercially available from Clontech (Palo Alto, CA). Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one includes the DNA-binding domain of a transcription activator protein fused to one test protein "X" and the other includes the activator protein's activation domain fused to another test protein "Y". Thus, either "X" or "Y" in this system
5 can be a host protein (such as those listed in Table 1 and any target sequences associated with SEQ ID NOS: 1-232), while the other can be a test protein or peptide. The plasmids are transformed into a strain of *Saccharomyces cerevisiae* that contains a reporter gene (such as lacZ) whose regulatory region contains the activator's binding sites. Either hybrid protein alone cannot activate transcription of the reporter gene, the DNA-binding domain hybrid because it does not provide activation function
10 and the activation domain hybrid because it cannot localize to the activator's binding sites.
Interaction of the two proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

The two-hybrid system or related methodology can be used to screen activation domain libraries for proteins that interact with a host protein involved in viral infection. Total genomic or
15 cDNA sequences are fused to the DNA encoding an activation domain. This library and a plasmid encoding a hybrid of the host protein involved in viral infection fused to the DNA-binding domain are cotransformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene. These colonies are purified and the plasmids responsible for reporter gene expression are isolated. DNA sequencing is then used to identify the proteins encoded by the
20 library plasmids.

For example, and not by way of limitation, a host gene encoding a protein involved in viral infection (such as those listed in Table 1 and target sequences associated with SEQ ID NOS: 1-232) can be cloned into a vector such that it is translationally fused to the DNA encoding the DNA-binding domain of the GAL4 protein. A cDNA library of the cell line from which proteins that interact with
25 the host protein are to be detected can be made using methods routinely practiced in the art. In this particular system, the cDNA fragments can be inserted into a vector such that they are translationally fused to the activation domain of GAL4. This library can be co-transformed along with the host-GAL4 DNA binding domain fusion plasmid into a yeast strain which contains a lacZ gene driven by a promoter which contains GAL4 activation sequences. A cDNA encoded protein, fused to GAL4
30 activation domain, that interacts with the host protein will reconstitute an active GAL4 protein and thereby drive expression of the lacZ gene. Colonies which express lacZ can be detected by their blue color in the presence of X-gal. The cDNA can then be extracted from strains derived from these and used to produce and isolate the host protein-interacting protein using techniques routinely practiced in the art.

35

Example 12

Rapid Screening Assays

Prior to performing any assays to detect interference with the association of a host protein involved in viral infection and a viral protein such as an HIV, Ebola, or influenza A protein, rapid

screening assays can be used to screen a large number of agents to determine if they bind to the host or viral protein. Rapid screening assays for detecting binding to HIV proteins have been disclosed, for example in U.S. Patent No. 5,230,998, which is incorporated by reference. In that assay, a host protein (such as those listed in Table 1 and target protein sequences associated with SEQ ID NOS: 1-232) or a viral protein, such as an HIV protein, is incubated with a first antibody capable of binding to the host or viral protein, and the agent to be screened. Excess unbound first antibody is washed and removed, and antibody bound to the host or viral protein is detected by adding a second labeled antibody which binds the first antibody. Excess unbound second antibody is then removed, and the amount of the label is quantitated. The effect of the binding effect is then determined in percentages by the formula: (quantity of the label in the absence of the test agent) - (quantity of the label in the presence of the test agent / quantity of the label in the absence of the test agent) x 100.

Agents that are found to have a high binding affinity to the host or viral protein can then be used in other assays more specifically designed to test inhibition of the host protein/viral protein interaction, or inhibition of viral replication.

15

Example 13

Assays for Measuring Inhibition of Viral Infection

Any of the test agents identified in the foregoing assay systems can be tested for their ability to decrease or inhibit infection by a pathogen or virus such as HIV, Ebola, or influenza A.

20

Cell-based assays

Exemplary methods are provided in Example 3 above. Briefly, cells (20,000 to 250,000) are infected with the desired pathogen, such as HIV, Ebola, or influenza A, and the incubation continued for 3-7 days. The test agent can be applied to the cells before, during, or after infection with the virus. The amount of virus and agent administered can be determined by skilled practitioners. In some examples, several different doses of the potential therapeutic agent can be administered, to identify optimal dose ranges. Following transfection, assays are conducted to determine the resistance of the cells to infection by various agents.

For example, the presence of a viral antigen can be determined by using antibody specific for the viral protein then detecting the antibody. In one example, the antibody that specifically binds to the viral protein is labeled, for example with a detectable marker such as a fluorophore. In another example, the antibody is detected by using a secondary antibody containing a label. The presence of bound antibody is then detected, for example using microscopy, flow cytometry, and ELISA.

Alternatively or in addition, the ability of the cells to survive viral infection is determined, for example by performing a cell viability assay, such as trypan blue exclusion.

Animal model assays

The ability of an agent, such as those identified using the methods provided above, to prevent or decrease infection by a virus, such as HIV, Ebola, or influenza A, can be assessed in animal

models. Several animal models for viral infection are known in the art. For example, mouse HIV models are disclosed in Sutton *et al.* (*Res. Initiat Treat. Action*, 8:22-4, 2003) and Pincus *et al.* (*AIDS Res. Hum. Retroviruses* 19:901-8, 2003); guinea pig models for Ebola infection are disclosed in Parren *et al.* (*J. Virol.* 76:6408-12, 2002) and Xu *et al.* (*Nat. Med.* 4:37-42, 1998); and cynomolgus monkey (*Macaca fascicularis*) models for influenza infection are disclosed in Kuiken *et al.* (*Vet. Pathol.* 40:304-10, 2003). Such animal models can also be used to test agents for an ability to ameliorate symptoms associated with viral infection. In addition, such animal models can be used to determine the LD50 and the ED50 in animal subjects, and such data can be used to determine the *in vivo* efficacy of potential agents.

10 Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, such as baboons, monkeys, and chimpanzees, can be used to generate an animal model of viral infection if needed.

15 The appropriate animal is inoculated with the desired virus, in the presence or absence of the test agents identified in the examples above. The amount of virus and agent administered can be determined by skilled practitioners. In some examples, several different doses of the potential therapeutic agent can be administered to different test subjects, to identify optimal dose ranges. The therapeutic agent can be administered before, during, or after infection with the virus. Subsequent to the treatment, animals are observed for the development of the appropriate viral infection and symptoms associated therewith. A decrease in the development of the appropriate viral infection, or 20 symptoms associated therewith, in the presence of the test agent provides evidence that the test agent is a therapeutic agent that can be used to decrease or even inhibit viral infection in a subject.

25 Having illustrated and described the principles of the invention by several examples, it should be apparent that those embodiments can be modified in arrangement and detail without departing from the principles of the invention. Thus, the invention includes all such embodiments and variations thereof, and their equivalents.